HARVARD
Journal on Legislation

Volume 39, Number 1 Winter, 2002

Articles

1 The Forgetfulness of Noblesse: A Critique of the German Foundation Law Compensating Slave and Forced Laborers of the Third Reich
Libby Adler & Peer Zumbansen

63 On the Economic Efficiency of Using Law to Increase Research and Development: A Critique of Various Tax, Antitrust, Intellectual Property, and Tort Law Rules and Policy Proposals
Richard S. Markovits

Symposium: Reforming Punitive Damages

121 Erratic by Design: A Task Analysis of Punitive Damages Assessment
David A. Schkade

139 Punitive Damages: How Jurors Fail to Promote Efficiency
W. Kip Viscusi

Note

169 State Employees and Sovereign Immunity: Alternatives and Strategies for Enforcing Federal Employment Laws
Brent W. Landau

Recent Developments

215 The Unborn Victims of Violence Act

237 The Bankruptcy Abuse Prevention and Consumer Protection Act

255 Mental Health Parity

Harvard Journal on Legislation
Publications Center
Harvard Law School
1541 Massachusetts Avenue
Cambridge, MA 02138
(617) 495-4400
hlsjol@law.harvard.edu
www.law.harvard.edu/studorgs/jol

Editorial Policy: The Journal specializes in the analysis of legislation and the legislative process. It focuses on legislative reform and on organizational and procedural factors that affect the efficiency and effectiveness of legislative decision-making. The Journal is especially interested in publishing articles that examine public policy problems of nationwide significance and propose legislation to resolve them.

Submissions: The Journal welcomes unsolicited manuscripts. Interested authors should send their manuscripts and all related correspondence to the Articles Editor at the above address. Submissions by mail should be typed, double-spaced, and accompanied by a diskette containing the submission in Microsoft Word. Submissions by e-mail should be attached as a Microsoft Word document, and the text of the e-mail should include the author's complete contact information. Footnotes should comply with *The Bluebook: A Uniform System of Citation* (17th ed. 2000).

Subscriptions: Subscription rates for the 2002 volume (Volume 39) are $28.00 domestic; $34.00 foreign surface mail ($36.00 to Canada); $46.00 foreign airmail. Individual issues from the current volume can be obtained for $14.00 (plus $3.00 per copy for foreign mailing). All correspondence concerning subscriptions should be addressed to the Subscriptions Manager, Publications Center, Harvard Law School, 1541 Massachusetts Avenue, Cambridge, MA 02138, (617) 495-3694. E-mail: ssulliva@law.harvard.edu. Unless otherwise requested by the subscriber, the *Journal* is mailed at bulk rate. For some overseas subscribers, delivery via air mail may be more reliable than standard surface rate service. Claims for missing issues of Volume 39 will be honored only until January 2003.

Back Issues: Past volumes are available exclusively from William S. Hein & Co., Inc., 1285 Main St., Buffalo, NY, 14209, 1-800-828-7571, wsheinco@class.org. Contact William S. Hein & Co. for information on prices for bound volumes.

ARTICLE

THE FORGETFULNESS OF NOBLESSE: A CRITIQUE OF THE GERMAN FOUNDATION LAW COMPENSATING SLAVE AND FORCED LABORERS OF THE THIRD REICH

LIBBY ADLER*
PEER ZUMBANSEN**

This Article analyzes the Law on the Creation of the Foundation "Remembrance, Responsibility and Future" which the German Legislature passed on July 17, 2000. The law established a foundation designed to compensate individuals who were victims of the Nazi slave and forced labor program. Because of their roles in this atrocity, the German government and German industry both contributed billions of dollars to the Foundation. In this Article, the authors identify numerous flaws with the law. They argue that, despite the monetary payments, no entity took true responsibility for the forced labor program. Various groups were excluded from the negotitations and some had their claims extinguished by the law but were not afforded compensation by it. The authors also contend that much of the justification for the Foundation Law originates in flawed and unexamined assumptions about the strength of the victims' legal claims. These factors, and others, have led the authors to conclude that this law fails to achieve its primary goal, the remembrance of the horrific acts committed by the Nazis.

I. INTRODUCTION

In August 2000, Germany's twin houses of parliament enacted a law establishing a foundation to compensate survivors of the Nazi forced la-

---

* Assistant Professor of Law, Northeastern University; A.B., University of Michigan, 1989; J.D., Northeastern University School of Law, 1994; LL.M., Harvard Law School, 1998. I am grateful to Nathaniel Berman, Karl Klare, Wendy Parmet, and Lucy Williams for their helpful comments on earlier drafts and to Brian Birke and Roy Karp for their research assistance. I also wish to express appreciation to my friend, colleague, and co-author, Peer Zumbansen. It will be years before I complete the inventory of all that I have learned from our time working together. My contribution to this Article is dedicated to my grandfather, Ed Adler.

** Academic Assistant and Lecturer, University of Frankfurt, Institute of Economic Law; Jean Monnet Fellow, European University Institute, Florence, Italy; Licence en Droit, Université Paris-X-Nanterre, 1991; 1st German Law Exam 1995; Dr. iur., University of Frankfurt, 1998; LL.M., Harvard Law School, 1998; 2d German Law Exam 2000. My heartfelt gratitude goes to Günter Frankenberg, Alexander Hanebeck, Dominik Hanf, and to Libby Adler, who has been a friend and companion, and who, during both the joint seminar at the University of Frankfurt in December 2000 and during the work for this Article, has been an overwhelming source of encouragement and inspiration.

bor program.[1] The Foundation Law has been acclaimed as a victory for Holocaust survivors. In this Article, we examine the recent events and post-war historical developments that led to the Foundation Law and conclude that the victory is ambiguous at best.

While the Foundation Law provides compensation to Nazi-era slave and forced laborers, it also sharply limits compensation amounts, denies recovery to some potential claimants, and purports to preclude further litigation of Holocaust-era claims. Proponents of the Foundation Law have defended the choice to use legislation to resolve Holocaust-related claims initially brought in a judicial forum on the grounds that litigation is inherently ill-suited to that task, and have justified the terms of the Law by reference to claimants' poor chances in the courtroom. In this Article, we identify some troubling assumptions underlying these rationales and highlight the historical and political context in which they are offered.

Germany enacted the Foundation Law pursuant to an international agreement designed to resolve legal claims against German companies arising out of their use of slave and forced labor during the Third Reich.[2] In accordance with the Law, *slave laborers*, including Jews, Sinti and Roma (or Gypsies), gay men and lesbians, and others whom the Nazis incarcerated in concentration camps and who worked under extraordinarily inhumane conditions,[3] are eligible to receive up to DM 15,000 each.[4] *Forced laborers*, including Poles, Czechs, and other nationals, who worked under slightly less brutal conditions and varying degrees of coercion for no (or extremely low) wages,[5] are eligible to receive up to DM 5,000 each.[6] Awards available under the Law are capped at these maximums. For the most part, heirs are not eligible to recover.[7]

---

[1] Gesetz zur Errichtung einer Stiftung "Erinnerung, Verantwortung, Zukunft" [Law on the Creation of the Foundation "Remembrance, Responsibility and Future"], v. 2.8.2000 (BGBl. I S.1263), *available in English at* http://www.stiftungsinitiative.de/eindex.html; *see generally* Hugo J. Hahn, *Individualansprüche auf Wiedergutmachung von Zwangsarbeit im Zweiten Weltkrieg*, 53 NEUE JURISTISCHE WOCHENSCHRIFT 3521 (2000); Christoph Safferling, *Zwangsarbeiterentschädigungsgesetz und Grundgesetz. Zur Frage der Verfassungsmäßigkeit des Gesetzes zur Errichtung einer Stiftung "Erinnerung, Verantwortung, Zukunft,"* 34 KRITISCHE JUSTIZ 208 (2001).

The Law also deals with insurance, property and personal injury claims, *see* Law on the Creation of the Foundation § 9(3)–(4), and creates a fund for educational and humanitarian projects, *see id.* §§ 2, 9(7).

[2] *Agreement Between the Government of the United States and the Government of the Federal Republic of Germany Concerning the Foundation "Remembrance, Responsibility, and Future"* (July 17, 2000) [hereinafter Agreement], *at* http://www.state.gov/www/regions/eur/holocaust/000717_agreement.html.

[3] *See* ULRICH HERBERT, A HISTORY OF FOREIGN LABOR IN GERMANY, 1880–1980 176 (1990).

[4] Law on the Creation of the Foundation, §§ 9(1), (11)1.

[5] *See* HERBERT, *supra* note 3, at 142.

[6] Law on the Creation of the Foundation, §§ 9(1), (11)1. At an approximate exchange rate of DM 2 to $1, claimants will receive close to $7,500 and $2,500, respectively.

[7] *Id.* § 13(1).

The Foundation itself was funded jointly by public and private entities. The German government and a consortium of German companies each committed DM 5 billion to the fund.[8] Non-governmental agencies, referred to in the Law as *partner organizations*, are cooperating to distribute this money to approximately one million people[9] residing primarily in the United States, Israel, Poland, the Czech Republic, Ukraine, Belarus, and other former Soviet republics.[10] These partner organizations are expected to publicize the work of the Foundation[11] as well as to evaluate claims and convey payment.[12] Judicial review of adverse determinations by the partner organizations appears to be precluded.[13]

This arrangement is the product of intense diplomatic negotiation in which officials of the United States as well as representatives of German industry and government played the most important roles.[14] The diplomatic efforts came on the heels of a slew of lawsuits filed in United States courts against German industry, European and American insurance companies, and Swiss banks, all of which allegedly acted tortiously or in breach of contract to the detriment of victims of Nazism or their heirs.[15]

Although litigation spurred the negotiations that led to the enactment of the Foundation Law, litigation ultimately was displaced by other legal practices. Since passage of the Foundation Law, virtually all lawsuits pending in United States courts involving claims arising out of slave or forced labor during the Third Reich have been dismissed. German courts, including the German Federal Court of Justice (*Bundesgerichtshof*) and the Federal Constitutional Court (*Bundesverfassungsgericht*), have dis-

---

[8] William Drozdiak, *Germans Reach Settlement with Slave Laborers*, WASH. POST, Dec. 18, 1999, at A20. This is not the first time German companies were called upon to compensate Jewish slave laborers. *See generally* BENJAMIN B. FERENCZ, LESS THAN SLAVES: JEWISH FORCED LABOR AND THE QUEST FOR COMPENSATION (1979). In the years following World War II, a handful of the hundreds of German companies that used concentration camp labor turned over a total sum of approximately $13 million, which was shared by fewer than 15,000 of the Jewish survivors. Many received less than $1,000. *Id.* at 188.

[9] Estimates range from 700,000 to 2.3 million, but the one million estimate is cited most frequently. *See* Michael J. Bazyler, *Nuremberg in America: Litigating the Holocaust in United States Courts*, 34 U. RICH. L. REV. 1, 192 (2000).

[10] *See* Law on the Creation of the Foundation § 9.

[11] *See id.*, § 10(2).

[12] *Id.* § 10.

[13] *Id.* § 19.

[14] For a recent account, see Lothar Evers, *Verhandlungen konnte man das eigentlich nicht nennen . . .*, *in* STIFTEN GEHEN: NS-ZWANGSARBEIT UND ENTSCHÄDIGUNGSDEBATTE 222–34 (Ulrike Winkler ed., 2000).

[15] *See* Burger-Fischer v. Degussa AG, 65 F. Supp. 2d 248 (D.N.J. 1999); Iwanowa v. Ford Motor Co., 67 F. Supp. 2d 424 (D.N.J. 1999). *See generally* Bazyler, *supra* note 9; Derek Brown, *Litigating the Holocaust: A Consistent Theory In Tort For The Private Enforcement Of Human Rights Violations*, 27 PEPP. L. REV. 553 (2000); Stuart M. Kreindler, *History's Accounting: Liability Issues Surrounding German Companies For The Use Of Slave Labor By Their Corporate Forefathers*, 18 DICK. J. INT'L. L. 343 (2000).

missed lawsuits filed in Germany, holding that the Law precludes claims of this nature.[16]

This is because in exchange for the *quid* of a commitment to create and fund the Foundation, Germany sought only one *quo*: *legal peace*, i.e., an end to the lawsuits. For this reason, the United States agreed to minimize the legal threat to German industry by filing a *Statement of Interest* in every American lawsuit involving a claim arising out of slave or forced labor during the Third Reich. The Statement of Interest calls for dismissal of the suit on political question or other grounds and states that it is in the national interest of the United States that the Foundation be the sole avenue for former slave and forced laborers to obtain compensation.[17]

The Law provides for the German parliament (*Bundestag*) to declare that legal peace has been achieved once the last lawsuit pending before a United States court is dismissed. This declaration was a prerequisite for the compensation provisions of the Law to become effective.[18] The Bundestag made this declaration on May 30, 2001.[19]

Proponents of the Law, such as former United States Deputy Treasury Secretary Stuart Eizenstat, who devoted tremendous effort to securing restitution for the survivors of Nazism,[20] offer various reasons why the United States has obliged itself to intervene in any future lawsuits on the side of German industry: Litigation can be costly and time-consuming.[21] The survivor population is rapidly aging and dying at a rate of one percent per month.[22] Moreover, German industry has strong legal defenses; even plaintiffs who live long enough to see the outcome of their suits are unlikely to prevail.[23] For these reasons, proponents of the Law urge, it is most prudent to curtail access to judicial remedies and redirect

---

[16] *See, e.g.*, BGH [Federal Court of Justice], NEUE JURISTISCHE WOCHENSCHRIFT ("NJW"), 54 (2001), 1069 (citing Law on the Creation of the Foundation § 16), *cert. denied*, Bundesverfassungsgericht ("BVerfG") [Federal Constitutional Court], NJW, 54 (2001), 2159; OLG Northrhine-Westfalia, NJW 53 (2000), 3577.

[17] *See* Agreement, *supra* note 2, Annex B.

[18] According to the Foundation Law, "[t]he first allocation of funds to the Foundation requires as a precondition the entry into force of the German-American Intergovernmental Agreement Concerning the Foundation 'Remembrance, Responsibility and Future,' and the establishment of adequate legal security for German enterprises. The German Bundestag shall determine whether these preconditions exist." Law on the Creation of the Foundation § 17(2).

[19] Roger Cohen, *Last Chapter: Berlin to Pay Slave Workers Held by Nazis*, N.Y. TIMES, May 31, 2001 at A7.

[20] *See* Stuart Eizenstat, *Justice for the Survivors*, WASH. POST, Jan. 16, 2001 at A21.

[21] *Eizenstat Testimony on Worldwide Holocaust Restitution Efforts: Hearing before the Senate Foreign Relations Comm.*, 106th Cong. (Apr. 5, 2000) (statement of Eizenstat, U.S. Deputy Sec'y and Special Rep. for Holocaust Issues).

[22] *Id.*

[23] *Id.* These arguments echo the position of corporate defendants in past suits for compensation and are bolstered by the affidavits of sympathetic academics attesting to a total lack of corporate legal responsibility. *See* RUDOLF RANDELZHOFER & OLIVER DÖRR, ENTSCHÄDIGUNG FÜR NS-ZWANGSARBEIT, 1994 (affidavit prepared for German government in case of forced labor compensation before the BVerfG in 1996).

claimants to the Foundation, which promises expeditious processing and disbursement. Private distribution, proponents declare, is designed to ensure the speediest, least bureaucratic avenue to satisfying valid claims.[24]

Supporters' arguments, however, go beyond mere pragmatism. Proponents often argue that the claims for compensation were "not about the money"; it was *memory* or *acknowledgement* that claimants really sought.[25] To such a hallowed purpose, litigation—the site of objections, cross-examinations, delay tactics, and greedy lawyers—appears ill-suited, even indecent. Visualize instead an affirmative legislative act agreed to through international cooperation and accompanied by a solemn profession of a moral obligation to those who have awaited justice for so many years.[26] When litigation is juxtaposed against this noble alternative, for a cause as sacrosanct as compensating the survivors of Nazism, the proper course seems clear.

On the other hand, this course also enabled German industry to avert the assignment of legal responsibility for the enslavement of millions of civilian workers from 1933 to 1945.[27] From the outset, the companies consistently and sternly denied any legal culpability, preferring instead to assume a "moral obligation" to compensate the survivors of their predecessors' brutality.

In this Article, we do not establish that the continuation of lawsuits under the current American class action regime would have generated a larger financial settlement. Our thesis concerns the Law's pretense to providing not only material restitution to survivors of the Nazi labor program, but also acknowledgement of responsibility for the atrocities that the survivors endured. We contend that a close look at the Law belies any real acknowledgement of responsibility by German industry. A close look reveals no remorse, no confession, and no sense of debt for the mer-

---

[24] *See, e.g.*, Stuart Eizenstat White House briefing, Dec. 15, 1999.

[25] *See* Evers, *supra* note 14, at 225–26 (quoting Volkswagen's Head of Media Relations as saying, "No, I really think that we ought to proceed in encouraging and funding political publicity work under the Heading: 'Never again Fascism . . .' We should not go the way of individual compensation payments and backroom agreements. I don't even think that there is an adequate sum—and I am repeating it again, for us this is not about money." (transl. by Zumbansen)).

[26] The Foundation Law's Preamble contemplates historical, political, and moral responsibility. Law on the Creation of the Foundation, Preamble, v. 2.8.2000 (BGBl. I S.1263).

[27] Recently, the responsibility of governmental and private actors for slavery during Hitler's reign has been widely researched, in the academy as well as by the corporations themselves. *See, e.g.*, BARBARA HOPMANN ET AL., ZWANGSARBEIT BEI DAIMLER-BENZ (1994); HANS MOMMSEN & MANFRED GRIEGER, DAS VOLKSWAGENWERK UND SEINE ARBEITER IM DRITTEN REICH (1996); ZWANGSARBEIT BEI FORD (Projektgruppe "Messelager" im Verein EL-DE-Haus e.V. Köln ed. 1996); I. G. FARBEN, VON ANILIN BIS ZWANGSARBEIT (Coordination gegen BAYER-Gefahren e.V./CGB, eds., 1995); Bernd C. Wagner, IG AUSCHWITZ: ZWANGSARBEIT UND VERNICHTUNG VON HÄFTLINGEN DES LAGER MONOWITZ 1941–1945, Darstellungen und Quellen zur Geschichte von Auschwitz (Institut für Zeitgeschichte ed., 2000); Peter Hayes, *Zur umstrittenen Geschichte der I. G. Farbenindustrie AG*, 18 GESCHICHTE UND GESELLSCHAFT 405 (1992).

ciless treatment of these people who exerted themselves beyond human capacity. A close look reveals not a crack in the companies' unwavering avowal of their blamelessness.

It may be true that some claims against the German companies were "not about the money," but from this it does not necessarily follow that any settlement suffices. If acknowledgement or memory was the claimants' primary objective, it is at least worthwhile to ask whether the Law achieved it. This Article examines the Law, the process by which agreement on its terms was reached, the promises exchanged for it, and the rationales offered in its support, in an effort to address that question. We argue that with the aid of German and American courts, the German companies made a series of shrewd institutional maneuvers that quietly defeated claimants' acknowledgement objective, again and again avoiding confrontation with their pasts. Furthermore, the companies and their partners in negotiation bolstered support for these moves, perversely, by reference to the sanctity of the memory of the Holocaust and its incompatibility with money damages and litigation. In the end, while the Foundation Law was touted as achieving memory, it achieved forgetting.

It may be too much to expect a piece of legislation to perform the comprehensive work of coming to terms with the past, but one can at least be skeptical about representations that credit the Foundation Law for having achieved such an admirable goal. As we will discuss, claims alleging specific corporate culpability have consistently been channeled away from concrete, plaintiff-versus-defendant confrontations, toward more abstract declarations of "moral," "historical," or "political" responsibility. We attempt to demonstrate that the Law is a recent development in a trend reaching back to the post-war years. Since World War II, redress for victims of the Holocaust has been treated as a matter of foreign affairs and international reparations rather than litigation between private parties. This preference for extra-judicial remedies originated in a desire to enable German industry to recuperate economically following World War II. Now, with German industry long since recovered, the notion that litigation is an inappropriate means of addressing Holocaust victims' concerns has persisted and lawsuits for individual relief have met largely with failure.

Consistent with its history, the Law conditions the disbursement of compensation payments on legal peace and purports to limit all future claimants to its terms. Moreover, the preamble acknowledges the propriety of compensating victims, but does not specifically assign responsibility for perpetrating the offenses that gave rise to the need for compensation. Literature on the topic of personal and collective responsibility for historical misdeeds is not scant,[28] but the Foundation Law reflects none of this grappling; this is what troubles us.

---

[28] *See, e.g.*, THEODOR W. ADORNO, *The Meaning of Working Through the Past*, *in*

Recently, other racial and national groups that have endured historical atrocities have come forward with claims against governments and private entities that participated in their subjugation. African American scholars and activists have commenced a national debate on restitution for American slavery.[29] Women from the Pacific who were forced to serve as "comfort women" to the Japanese military in the 1930s and 1940s seek restitution from the government of Japan.[30] On occasion, spokespersons for these efforts cite the German Foundation as an example of the kind of restitution they seek, admiring the success that victims of Nazism had in obtaining compensation and acknowledgement for the wrongs they suffered.[31] We hope that the Foundation will serve as an example to these other claimants, but not as a model to be emulated—rather as a cautionary tale to be heeded.

Part II of this Article provides an account of the steps that led to enactment and implementation of the Law, from the initiation of lawsuits in United States courts in the 1990s, through transatlantic negotiations, to the Bundestag's clearing the way for payments to flow. Part III places the Law in historical context, showing how German and American courts repeatedly have blocked private claims, describing the claims in terms that render them best resolved by international diplomacy. This Part also unearths the tangle of political and moral judgments underlying the persistent denial of corporate legal responsibility. Part IV examines the rhetorical strategy that proponents deployed to bolster the legitimacy of the Foundation Law. Proponents spoke of the Holocaust as if it were uniquely incompatible with litigation. In so doing, they created a hostile climate for legal claims and diverted attention to an acknowledgement objective that was never achieved. Finally, Part V concludes that the Foundation Law, far from being an unambiguous success for Holocaust survivors, defeated memory and instead produced forgetfulness.

CRITICAL MODELS: INTERVENTIONS AND CATCHWORDS 89–103 (Henry W. Pickford trans., 1998) (1963); ALEXANDER & MARGARETE MITSCHERLICH, THE INABILITY TO MOURN: PRINCIPLES OF COLLECTIVE BEHAVIOR (Beverley R. Placzek trans., 1975) (1967); CHRISTIAN MEIER, VIERZIG JAHRE NACH AUSCHWITZ (1990); JÜRGEN HABERMAS, *Was Bedeutet »Aufarbeitung der Vergangenheit« Heute? in* DIE NORMALITÄT EINER BERLINER REPUBLIK 21–45 (1995); PETER NOVICK, THE HOLOCAUST IN AMERICAN LIFE (1999); CHARLES MAIER, THE UNMASTERABLE PAST. HISTORY, HOLOCAUST AND GERMAN NATIONAL IDENTITY (1988).

[29] *E.g.*, RANDALL ROBINSON, THE DEBT: WHAT AMERICA OWES TO BLACKS (2000).

[30] *All Things Considered: Comfort Women* (NPR broadcast, Aug. 1, 2001).

[31] *See id.*

II. Background of the Foundation Law: The Move from Litigation to Diplomacy and Legislation

A. American Lawsuits in the 1990s

In the 1990s,[32] Holocaust survivors began filing lawsuits in United States courts against Germany and German companies for enslavement during the Third Reich.[33] One such plaintiff was Hugo Princz.[34] At age 16,[35] Princz and his family, all American citizens, found themselves trapped in Czechoslovakia when the United States declared war against Germany in 1942.[36] All six were arrested as enemy aliens and sent to concentration camps.[37] Princz's parents and sister were executed at Treblinka.[38] Princz and his brothers were sent to Birkenau, where they performed slave labor at a chemical plant owned by I. G. Farben.[39] Only Princz survived.[40] He was subsequently transferred to Dachau, where he performed slave labor at an underground airplane factory,[41] and was *en route* to a third concentration camp for extermination when, at the end of the war, the United States liberated the freight train in which Princz was being transported.[42]

Nearly a half-century later, after years of failed attempts to obtain payments directly from the German government,[43] Princz sued Germany

---

[32] Litigation arising out of the Holocaust began much earlier, but the lawsuits initiated between the 1950s and 1980s received little public or academic attention. *See* Walter Schwarz, *Abschied*, Rechtsprechung zum Wiedergutmachungsrecht (RzW), 32 (1981), 114 (discussing suits from 1950 to 1981).

[33] These lawsuits were spurred in part by the conclusion of the Cold War and the reunification of Germany, both of which resulted in the declassification of documents previously unavailable to plaintiffs. *See* Allyn Z. Lite, *Another Attempt to Heal the Wounds of the Holocaust*, 27 SPG Hum. Rts. 12, 14 (2000).

[34] Princz v. Federal Republic of Germany, 26 F.3d 1166 (D.C. Cir. 1994).

[35] *Id.* at 1176 (Wald, J., dissenting).

[36] *Id.* at 1168.

[37] *Id.*

[38] *Id.*

[39] *Id.* I. G. Farben, now in liquidation, is a large chemical concern composed of several smaller companies; the most well-known in the United States is probably Bayer. *See* Joseph Borkin, The Crime and Punishment of I. G. Farben 4, 160–63 (1978). Its notorious Buna IV plant exploited the slave labor of Auschwitz-Birkenau inmates in the production of *buna*, a synthetic rubber that Hitler hoped would help sustain his war effort. *See id.* at 111–27.

[40] *See Princz*, 26 F.3d at 1168.

[41] *Id.* at 1177 (Wald, J., dissenting).

[42] *Id.* at 1168.

[43] *Id.* at 1177 (Wald, J., dissenting). Princz was deemed ineligible for an existing compensation program operated by the German government because he was a citizen of the United States at the time of his enslavement. *Id.* The program, known as the Bundesentschädigungsgesetz ("BEG"), enacted on June 26, 1956, (BGBl. I S.559), and amended on September 14, 1965, (BGBl. I S.1315), provides compensation to those who were persecuted by the Nazis for political, racial, religious or ideological reasons, but does not provide specific compensation for forced laborers. *See* Wiedergutmachung und Kriegsfolgenliquidation. Geschichte-Regelungen-Zahlungen 104–12 (Hermann-Josef

in federal district court in Washington, D.C., seeking to recover damages for injuries arising out of his enslavement.[44] The German government moved to dismiss Princz's complaint on the grounds that the Foreign Sovereign Immunities Act of 1976 ("FSIA")[45] immunized Germany against the claims asserted.[46] On appeal from the district court's denial of the motion,[47] the D.C. Circuit dismissed the case, pursuant to the FSIA, for lack of subject matter jurisdiction.[48]

Other former slaves took a cue from Princz's failed efforts at suing the German government and sought to avoid the FSIA by naming as defendants the private companies that profited from the use of slave labor. Former slaves of, among others, Ford Werke,[49] Degussa,[50] and Siemens[51]

Brodesser et al. eds., 2000) [hereinafter WIEDERGUTMACHUNG UND KRIEGSFOLGENLIQUIDATION]; *see* CORNELIUS PAWLITA, "WIEDERGUTMACHUNG" ALS RECHTSFRAGE? 301–07 (1993).

[44] *See Princz*, 26 F.3d at 1168.

[45] 28 U.S.C. §§ 1330, 1602–11.

[46] *See Princz*, 26 F.3d at 1168.

[47] 813 F. Supp. 22, 26 (D.D.C. 1992), *rev'd* 26 F.3d 1166 (D.C. Cir. 1994). The district court reasoned that the FSIA "has no role to play where the claims alleged involve undisputed acts of barbarism committed by a one-time outlaw nation which demonstrated callous disrespect for the humanity of an American citizen, simply because he was Jewish." *Princz*, 813 F. Supp. at 26. The Second Circuit, however, held that "that is not the law." *Princz*, 26 F.3d at 1169.

[48] *Princz*, 26 F.3d at 1168. After Princz's claims against the German government were dismissed, he attempted to bring essentially the same claims against four private companies, which he alleged to be the corporate successors to the companies that enslaved him. *See* Princz v. BASF Group, No. 92-0644, 1995 U.S. Dist. LEXIS 22104, at *1–*3 (D.D.C. Sep. 18, 1995). In addition to BASF Group, Princz sued Hoechst AG, Bayer Group, and Daimler-Benz AG. The first three were alleged to be the successors to I. G. Farben. *Id.* at *3. Again, Princz's claims were dismissed, this time by stipulation of the parties. *Id.* at *17. In his order dismissing the case, District Court Judge Stanley Sporkin cited not only the court of appeals' reversal of his decision in the first suit, but also the availability of "meritorious legal defenses," including lack of personal and subject matter jurisdiction, *id.* at *6, *16, and the existence of pending government-to-government negotiations intended to resolve the sorts of claims that Princz raised in a litigation context, *id.* at *4. In 1994, then President Clinton raised Princz's claim in a meeting with then German Chancellor Helmut Kohl, who reportedly declined to discuss the matter. *See Princz*, 26 F.3d at 1177 (Wald, J., dissenting). Princz finally obtained financial remuneration pursuant to an agreement between Germany and the United States called *The Hugo Princz Agreement of 1995*, awarding Princz and ten other United States citizens caught in the Third Reich $2.1 million. *See* Bazyler, *supra* note 9, at 25; Symposium, *What Happens Next?*, 20 WHITTIER L. REV. 91 (1998).

[49] *See* Iwanowa v. Ford Motor Co., 67 F. Supp. 2d 424 (D.N.J. 1999). Ford Werke AG is a wholly owned subsidiary of Ford, the American motor company. Ford established Ford Werke in 1925. *See id.* at 432. *See also* Michael Dobbs, *Ford and GM Scrutinized for Alleged Nazi Collaboration*, WASH. POST, Nov. 30, 1998, at A01 (discussing Ford's war-time engagement with Germany).

[50] *See* Burger-Fischer v. Degussa AG, 65 F. Supp. 2d 248 (D.N.J. 1999). Degussa (formerly Aktiengesellschaft Deutsche Gold- und Silber-Scheideanstalt, since 1999 Degussa-Hüls AG) is a large German chemical company that was incorporated in Frankfurt in 1873 and is a world leader in the development of exclusive metal applications as well as chemical products innovations. Degesch, a subsidiary of Degussa, developed and distributed Zyklon B, the gas used in the Nazi gas chambers. *See id.* at 252.

[51] *See id.* at 250 (naming Siemens as a co-defendant). Siemens AG, a large German

brought class action lawsuits against their former corporate masters in federal district court in the late 1990s.[52]

In addition to the prospect of substantial class action awards, the defendants feared that the lawsuits would damage their public image and thereby adversely affect German export industries.[53] The German corporate community faced criticism in the United States and elsewhere as the details of the forced labor program reached the public.[54] In early 1999, public attention in the United States reached its peak when New York officials asked regulators to delay a proposed merger between Deutsche Bank AG and Bankers Trust until the compensation issue was satisfactorily resolved.[55]

Around this time, respected art galleries were confronted with claims that valued items in their collections had been stolen during the "Aryanization"[56] of Jewish property in the 1930s.[57] Life insurance com-

---

close corporation founded in 1847, developed into a world leader in electric, electronic and technical equipment. Among Siemens's employees during the war were numerous slave and forced laborers from the Oranienburg, Buchenwald, Sachsenhausen, Flossenbürg and Auschwitz camps. *See* FERENCZ, *supra* note 8, at 117–22.

[52] *See, e.g.*, *Iwanowa*, 67 F. Supp. 2d at 424; *Degussa*, 65 F. Supp. 2d at 248; Class Action Complaint 1, Pollack v. Siemens AG, No. CV 98-5499 (E.D.N.Y., filed Aug. 30, 1998) (in which Krupp, Henkel, Audi, Daimler-Benz, Volkswagen, BMW, and other companies were also named as defendants); Fishel v. BASF, 175 F.R.D. 525 (S.D. Iowa 1997) (naming I. G. Farben and its successor companies as well as Daimler-Benz and Krupp as defendants). We discuss some of the issues in these cases in Part III below.

These companies were by no means the only entities to use slave labor during the Hitler years. Hundreds of companies (including Krupp, Daimler-Benz, and Volkswagen to name just a few) as well as farms, churches, and private households made widespread use of this cheap resource. *See generally* FERENCZ, *supra* note 8; Ulrich Herbert, *Labor as Spoils of Conquest, 1933–1945*, *in* NAZISM AND GERMAN SOCIETY 219–73 (David F. Crew ed., 1994); ULRICH HERBERT, FREMDARBEITER: POLITIK UND PRAXIS DES "AUSLÄNDER-EINSATZES" IN DER KRIEGSWIRTSCHAFT DES DRITTEN REICHES (2d ed. 1999) [hereinafter HERBERT, FREMDARBEITER]; Wolfgang Benz, *Zwangsarbeit im nationalsozialistischen Staat*, 16 DACHAUER HEFTE 3, 3–17 (2000); Katharina Hoffmann, *Lebensverhältnisse von ausländischen Zwangsarbeiterinnen und Zwangsarbeitern in der Stadt Oldenburg während des Zweiten Weltkriegs*, *in* NATIONALSOZIALISMUS UND ZWANGSARBEIT IN DER REGION OLDENBURG 79–122 (Katharina Hoffman & Andreas Lembeck eds., 1999).

[53] *See* Justin H. Roy, *Strengthening Human Rights Protection: Why the Holocaust Slave Labor Claims Should be Litigated*, 1 SCHOLAR: ST. MARY'S L. REV. MINORITY ISSUES 153, 155 (1999).

[54] *See, e.g.*, Ian Traynor, *Poles Seek Reparations for Forced Labor by Nazis*, DETROIT NEWS, Jan. 11, 1999, at A4.

[55] *See* Christopher Rhoades, *Bonn Airs Plan on Slave-Labor Claims: Fund, Foundation are Seen; Deutsche Bank's Deal in U.S. Could Benefit*, WALL ST. J., Feb. 10, 1999, at A19. In addition to having maintained accounts containing gold and proceeds from the sale of "Aryanized" property, Deutsche Bank disclosed in 1999 that it had helped finance the construction of Auschwitz. *See* Christopher Rhoades, *Deutsche Bank Discloses Involvement in the Financing of Nazi Death Camp*, WALL ST. J., Feb. 5, 1999, at A13.

[56] *See* Bazyler, *supra* note 9, at 96.

[57] *See* Stephan J. Schlegelmilch, *Ghosts of the Holocaust: Holocaust Victim Fine Arts Litigation and a Statutory Application of the Discovery Rule*, 50 CASE W. RES. L. REV. 87, 96–98 (1999) (generally describing recent claims regarding art stolen during World War II); Bazyler, *supra* note 9, at 161–90 (discussing the theft of art by the Nazis and recounting the details of some recent stolen art claims).

panies such as the Italian insurer *Generali*[58] faced claims by unpaid surviving beneficiaries.[59] Renewed allegations surfaced that Swiss banks also owed an overdue debt stemming from their concealment of records of deposits made by victims of Nazism, their collusion with the Nazi regime in laundering looted gold and other property, and their transactions in the profits of slave labor.[60] The insurers and Swiss banks were named in lawsuits that eventually resulted in substantial settlements.[61]

The late 1990s also saw a flurry of legislative activity at both state and federal levels concerning restitution for Nazi era crimes.[62] California, for example, enacted dramatic provisions in 1998 aimed at ensuring that California residents who are unpaid beneficiaries of Nazi era life insurance policies have an opportunity to recover under state law.[63] The following year, California statutorily granted jurisdiction to its courts to

---

[58] *See* Bazyler, *supra* note 9, at 97 (discussing Generali's founding and position in Europe).

[59] *See* Symposium, *supra* note 48, at 126. These companies denied their obligations to pay on such grounds as the claimant's failure to produce a death certificate for a decedent who had perished in a Nazi death camp, or that payments were already made to the Nazis as policy-holders. *See id.* at 126–27, 136–37 (1998).

[60] *See* Jeffrey Craig Mickletz, *An Analysis of the $1.25 Billion Settlement Between Swiss Banks and Holocaust Survivors and Holocaust Victims' Heirs*, 18 DICK. J. INT'L L. 199, 200 (1999). The banks claimed that they could not find the accounts or insisted that the heirs of the account holders produce death certificates, which of course could not be done. *See id.* at 203–04, 208. Initially, the Swiss banks also denied knowing the origins of the looted gold they received from the Nazis; they later abandoned this contention in light of evidence to the contrary. *See id.* at 207–08; *see also In re* Holocaust Victims' Assets Litigation, July 28, 2000 N.Y.L.J. 36 (col. 5).

[61] Generali, among other European insurers, was named in a federal class action filed in New York in 1997. *See* Amended Complaint, Connell v. Generali, No. 97-2262 (S.D.N.Y. filed June 26, 1997). Generali agreed to pay $100 million to compensate Holocaust era beneficiaries and make its records available for inspection. *See* Leonid Krechmer, *Holocaust-Related Claims and Limitations: Familiar Issues in a New Context*, 67 DEF. COUNS. J. 80, 83 (2000). Generali also was sued in a high profile case in state court in California that settled for an undisclosed amount. *See* Complaint, Stern v. Assicurazioni Generali S.p.A., No. BC 185376 (Cal. Super. Ct. filed Feb. 5, 1999); Bazyler, *supra* note 9, at 123–34; Symposium, *supra* note 48, at 136–38.

Four separate lawsuits were filed against Swiss banks and consolidated before Judge Korman of the Eastern District of New York, resulting in a $1.25 billion settlement approved July 28, 2000. *See* In re Holocaust Victims' Assets Litigation, July 28, 2000 N.Y.L.J. 36 (col. 5). For an analysis of the Swiss banks' settlement, see generally Mickletz, *supra* note 60.

German, Austrian and French banks also faced class action lawsuits in the United States for various war-time transgressions. *See* In re Austrian and German Bank Holocaust Litigation, 80 F. Supp. 2d 164 (S.D.N.Y. 2000); Complaint, Bodner v. Banque Paribas, No. CV-97-7433 (E.D.N.Y. filed Dec. 17, 1997).

[62] *See* Bazyler, *supra* note 9, at 272 app. B (listing "Federal and State Laws Regarding Holocaust Restitution").

[63] *See* CAL. CIV. PROC. CODE § 354.5 (entitling unpaid life insurance beneficiaries to bring a claim in California Superior Court until the year 2010); 1998 Cal. Legis. Serv. ch. 963, sec. 2 at 5564–65, codified in scattered sections of Cal. Ins. Code (requiring insurance companies to disclose information related to Nazi era insurance policies and creating a specialized team within the state's Department of Insurance to review insurance company archives). *See* Michael R. Santiago, *Victims No More: The War Over Holocaust Insurance Claims*, 30 MCGEORGE L. REV. 683, 684–88 (1999).

hear claims against corporations (and the successors of corporations) that used slave labor during the Nazi era and extended the statute of limitations for these claims to the year 2010.[64]

Amid this eruption of legal activity, representatives of German industry began holding meetings and soliciting affidavits from sympathetic experts as they deliberated on how to escape the American legal actions unscathed. The newly elected government of Germany, a center-left coalition of the Social Democrats and the Green Party, declared that compensation of slave laborers of the Third Reich would be one of its programmatic goals.[65] In October, 1998, less than a month after installation of the new coalition, members of the German corporate community asked their government to join them in establishing a foundation to compensate slaves of the Third Reich's industrial sector.[66]

---

[64] CAL. CIV. PROC. CODE § 354.6. *See generally* Diane Richard Foos, *Righting Past Wrongs or Interfering in International Relations? World War II-Era Slave Labor Victims Receive State Legal Standing After Fifty Years*, 31 MCGEORGE L. REV. 221 (2000); Russell A. Miller, *Much Ado, But Nothing: California's New World War II Slave Labor Statute of Limitations and Its Place in the Increasingly Futile Effort to Obtain Compensation From American Courts*, 23 WHITTIER L. REV. 121 (2001).

[65] *See* Bazyler, *supra* note 9, at 194. These parties, as well as the Free Democrats [FDP], had been proposing compensation for years, beginning with proposals by the Green Party, since their election into the Bundestag in 1983. They were joined by the Social Democrats in 1987 and 1989, but this and succeeding attempts never met with the Christian Democratic government's consent. *See* BT-Drs. 11/142 (parliamentary report issued on April 4, 1987, Otto Schily, Antje Vollmer, members of the Green Party and the German Bundestag, and the Green Parliamentary Group); BT-Drs. 11/4704 (parliamentary report issued on June 6, 1989, Antje Vollmer and the Green Parliamentary Group); BT-Drs. 11/4706 (parliamentary report issued on June 6, 1989, Antje Vollmer, Helmut Lippelt and the Green Parliamentary Group); BT-Drs. 11/5176 (parliamentary report issued on September 14, 1989, the Social Democratic Parliamentary Group). A public hearing in 1995 led to the provision of limited funds for certain groups of victims without providing for an encompassing regulation that would also benefit many of those persons now residing in Eastern European States. *See* Volker Beck, *Entschädigung für alle Verfolgten des Nationalsozialismus: Eröffnungsrede zur Anhörung der Bundestagsfraktion "Entschädigung für NS-Unrecht," in* ANERKENNUNG, REHABILITATION, ENTSCHÄDIGUNG: POLITISCHE INITIATIVEN FÜR DIE OPFER DES NATIONALSOZIALISMUS 50 JAHRE NACH KRIEGSENDE (Bündnis 90/Die Grünen im Bundestag et al., eds., 1995) (speech by Member of Parliament from the Green Party). Wiedergutmachung und Kriegsfolgenliquidation. Geschichte Regelungen Zahlungen (Hermann-Josef Brodesser et al. eds. 2000) at 54–55, listing agreements between Germany and Estland on 22 June 1995, Lithuania on 26 July 1996 and Lettland on 27 August 1998, leading to one-time payments of 2 Million DM each to be distributed to victims suffering from health damage incurred by Nazi persecution and now living in financial distress.

[66] It is difficult to discern which party initially proposed establishing a Foundation. The idea came up intermittently over the course of the 1980s and 1990s among participants in various lawsuits. *See supra* note 65 (citing Green Party proposals). Growing pressures on German industry in the late 1990s, mostly due to class action suits filed in United States courts, paved the way to passage of the Foundation Law. *See* Brief by Stuart Eizenstat (May 12, 1999) (reporting on the start of the negotiations and the participating parties), *at* http://www.ess.uwe.ac.uk/genocide/slave_labour2.htm (last visited Nov. 25, 2001).

For an account of individual settlements of the Jewish Claims Conference with German companies in the 1950s and 1960s, see FERENCZ, *supra* note 8. Ferencz himself proposed years ago that German industry create an "overall settlement" that would "diminish the prospects of a substantial number of lawsuits against many German firms." *Id.* at 181.

B. International Negotiations Regarding the Possibility of a Settlement

Litigation had drawn attention to the unsettled debt owed by German industry, but diplomatic negotiation would now entirely displace litigation as the institutional setting in which to resolve the claims.[67] In February 1999, Chancellor Gerhard Schröder announced a plan to establish a $1.7 billion foundation to be funded by a consortium of originally twelve, and quickly thereafter sixteen and then seventeen, German companies with a separate government fund to be established later.[68] Soon afterwards, representatives of Germany and German industry met with representatives of the United States and Jewish groups.[69] Attorneys for many of the United States plaintiffs participated in the meetings,[70] but some of the plaintiffs suing under California state law were denied permission to attend by the other participants.[71] The initial German offer of $1.7 billion was rejected, and negotiations progressed and faltered over the course of 1999, largely on the matter of the sum.[72]

On December 17, 1999, the parties agreed that the government and industrial sector of Germany would contribute DM 5 billion each, for a total of DM 10 billion or about $4.6 billion.[73] By this time, seventy companies had pledged to contribute.[74] In exchange for the fund, the United States agreed to oppose slave and forced labor lawsuits with a standard court filing stating that such suits are contrary to the national interests of the United States and should be dismissed.[75]

One industry representative responded that "a global approach was not considered feasible." *Id.* Another concluded "that any such agreement would be viewed as a confession of guilt by German industry, and efforts to pursue a global approach would only unite the companies in a solid front of opposition." *Id.* at 182.

[67] *See, e.g.*, D. Maimon, *Class Action Lawyers Charge German Slave Labor Fund Unacceptable*, *at* http://fido.seva.net:8090/pipermail/holocaust/1999-March/000042.html (last visited Nov. 25, 2001).

[68] *See* Bazyler, *supra* note 9, at 197 n.803; *see also* Burt Herman, *German Industry Raises Its Share of Nazi Labor Compensation Fund*, JERUSALEM POST, Mar. 14, 2001.

[69] *See id.* at 197.

[70] *See id.*

[71] Telephone Interview with Plaintiffs' Attorney Lisa Stern (Nov. 14, 2000). Stern speculated that she and her colleagues suing under California law were excluded from negotiations because the lawyers involved in the federal class actions had a much greater incentive to settle due to the perception that their cases were "in the toilet." *But see* Press Release, Lieff, Cabraser, Heimann & Bernstein (Dec. 15, 1999) (reporting on participation by some of the California lawyers), *at* http://www.lieffcabraser.com/slave_press.htm (last visited Nov. 25, 2001).

[72] *See* Bazyler, *supra* note 9, at 197–200.

[73] *See* William Drozdiak, *Germans Reach Settlement with Slave Laborers*, WASH. POST, Dec. 18, 1999, at A20.

[74] *See* Edmund L. Andrews, *Germany Accepts $5.1 Billion Accord to End Claims of Nazi Slave Workers*, N.Y. TIMES, Dec. 18, 1999, at A10. The contributions of the seventy committed companies were not expected to total the 5 billion DM that German industry was obligated to pay. *Id.*

[75] *See id.*

During the winter and spring of 2000, negotiations grew rancorous.[76] The principle issue of contention was how the money ought to be distributed among the survivors of the Nazi labor program, divided now into the categories of *slave laborers* and *forced laborers*.[77] *Slave laborers* encompass those workers who were imprisoned in concentration camps or detained in ghettos under extraordinarily brutal conditions.[78] *Forced laborers* include those who were deported from their home countries into the territory of the 1937 borders of the German Reich or to a German-occupied area;[79] they were subjected to forced labor in industry or for public authorities, and held in conditions less inhumane than a concentration camp.[80]

In March 2000, negotiators agreed that slave laborers would be eligible to receive up to DM 15,000 each, while forced laborers would be eligible for up to DM 5,000 each.[81] This disparity led to friction between Jews and non-Jewish, Central and Eastern European nationals.[82] Jews compose most of the population of concentration camp survivors and currently number in the hundreds of thousands.[83] Non-Jewish Central and Eastern Europeans compose most of the population of survivors of forced labor and number over a million.[84] Negotiations engendered further resentment among some Central and Eastern Europeans who complained about being excluded from the decision-making process and strong-armed by the United States into accepting the deal despite what they perceived to be low compensation amounts.[85]

---

[76] *See, e.g.*, Marilyn Henry, *German Firms Want Immunity from Israeli Suits in Slave-Labor Deal*, JERUSALEM POST, Mar. 5, 2000, at 3.

[77] *See id.*

[78] *See* Law on the Creation of the Foundation § 11(1), v. 2.8.2000 (BGBl. I S.1263).

[79] *See id.*

[80] *See, e.g.*, MARK SPOERER, ZWANGSARBEIT UNTER DEM HAKENKREUZ: AUSLÄNDISCHE ZIVILARBEITER, KRIEGSGEFANGENE UND HÄFTLINGE IM DEUTSCHEN REICH UND IM BESETZTEN EUROPA 1939–1945, 116–21 (2001) (describing the inhumane living conditions for forced laborers after 1942); HERBERT, FREMDARBEITER, *supra* note 52, at 170–73 (describing the high death rate of Russian forced laborers); *id.* at 334 (describing the difficult working and living conditions Russian forced laborers encountered due to malnutrition and lack of heating in work camps).

[81] *See* Law on the Creation of the Foundation § 9(1). In § 9(9), however, the Law limits initial payments to a percentage of the amounts set forth in § 9(1) until all applications have been processed, presumably in anticipation of discovering that the Foundation has insufficient funds to compensate everyone with a valid claim.

[82] *See* Andrews, *supra* note 74, at A10; Bazyler, *supra* note 9, at 201 n.814.

[83] *See* Andrews, *supra* note 74, at A10; Imre Karacs, *Germany's Pounds 3.2BN Bid to Close Book on Nazi Past Leaves only Rifts and Rancour in its Wake*, INDEP., July 18, 2000, at 3. The Nazi plan was to work Jewish camp inmates to death, and indeed only about five per cent of them survived. *See id.*; *see also* FERENCZ, *supra* note 8, at xvii (1979) (noting that Ferencz chose the title "Less Than Slaves" for his book "because our vocabulary has no precise word [for] unpaid workers who are earmarked for destruction.")

[84] *See* Andrews, *supra* note 74; Bazyler, *supra* note 9, at 201 n.814. Eighty-five percent of laborers in this category survived the war. *See* Karacs, *supra* note 83, at 3.

[85] *See id.*

In addition to the matter of distribution, negotiations hit a snag over the comprehensiveness of the *legal peace* to be guaranteed to Germany in exchange for the Foundation.[86] In an effort to advance the negotiations, the United States urged the governments of Israel and the Central and Eastern European states to promise German companies legal closure in their jurisdictions.[87] By May 2000, however, the discord focused on the degree of uncertainty inherent in the American system of separation of powers, according to which the executive branch could merely hope to persuade, but not require, a court to dismiss future lawsuits.[88]

On July 17, 2000, the Agreement was concluded.[89] Germany had committed itself to enacting legislation to create and fund the Foundation,[90] German industry had committed itself to contributing half of the money,[91] and the United States had committed itself to filing a *Statement of Interest* in all current and future cases in an effort to minimize the

---

[86] *See* Statement by Klaus Kohler, Chief, Deutsche Bank's Legal Department (Feb. 6, 2001) (stating that the provision of an "all-embracing and enduring legal peace" was the central objective of both the German industry representatives and the American and German governments participating at the negotiations), *at* http://www.spiegel.de/static/download/kohlerbrief.pdf (last visited Nov. 25, 2001).

[87] *See* Henry, *supra* note 76, at 3. The Agreement, while negotiated by the United States, is to be implemented through bilateral pacts in Eastern Europe. *See* Andrews, *supra* note 74, at A10. Some representatives from Eastern Europe were not happy with the deal negotiated by the United States but felt bullied into accepting it. *See* Karacs, *supra* note 83, at 3.

[88] *See* Richard Wolffe, *Clash of Cultures Threatens Holocaust Compensation Deal*, FIN. TIMES, May 27, 2000, at 6. While deliberating on the legality of President Carter's decision to negotiate a resolution to the Iranian crisis that curtailed court access for private litigants suing Iranian entities, however, then Justice Rehnquist observed that:

> the United States has repeatedly exercised its sovereign authority to settle the claims of its nationals against foreign countries. Though those settlements have sometimes been made by treaty, there has also been a longstanding practice of settling such claims by executive agreement without the advice and consent of the Senate. Under such agreements, the President has agreed to renounce or extinguish claims of United States nationals against foreign governments in return for lump-sum payments or the establishment of arbitration procedures.

Dames & Moore v. Regan, 453 U.S. 654, 679 (1981).

[89] *See* Edmund L. Andrews, *Germans Sign Agreement to Pay Forced Laborers of Nazi Era*, N.Y. TIMES, July 18, 2000, at A3.

[90] *See* Agreement, *supra* note 2, at Preamble ¶ 8.

[91] By this date, 3,127 companies had pledged to contribute, but only $1.6 billion had been raised. *See* William Drozdiak, *Germany Sets Fund For Slaves Of Nazis; $5 Billion Will Go To Aging Survivors*, WASH. POST, July 18, 2000, at A17. At least 200 German companies who had used slave and forced labor during Hitler's reign had yet to ante up. *See* Karacs, *supra* note 83, at 3. Though the Agreement does not explicitly say that funds will be contributed half by the German Government and the German industry, it provides that the Foundation be "funded by contributions from the Federal Republic of Germany and the German companies." This is in accordance with the wording of the Law. Law on the Creation of the Foundation § 3(2), v. 2.8.2000 (BGBl. I S.1263). After accounting for tax relief the companies will have been responsible for only about one-quarter of the Foundation's funding. *See* Jürgen Jeske, *Das lange Warten*, *in* FRANKFURTER ALLGEMEINE ZEITUNG No. 116, May 19, 2001, at 1.

threat to German industry in United States courts.[92] Despite objections to being excluded from key decisions, Poland, the Czech Republic, Russia, Belarus and the Ukraine also approved the Agreement.[93] Finally, Israel, the Conference on Jewish Material Claims Against Germany (the Jewish Claims Conference), and participating plaintiffs' attorneys also signed the deal.[94]

On August 2, 2000, Germany's Bundestag enacted the Law that established and funded the Foundation.[95] By its own terms, the Law took effect the same day.[96]

C. Complications in the Resolution of the American Lawsuits

Although the Law "took effect," the payments did not begin to flow. It would be almost a year before any of the survivors would begin to receive compensation.[97]

On January 24, 2001, plaintiffs in fifty class actions against German banks that had been consolidated before Judge Shirley Wohl Kram of the Southern District of New York asked the court to dismiss their claims in favor of recovery under the Foundation Law.[98] Despite this request, Judge Kram declined to dismiss the cases; instead, she took the matter under advisement, even though this decision meant postponing distribution of funds from the Foundation.[99] Representatives of the plaintiff class ex-

---

[92] *See* Agreement, *supra* note 2, at Art. 2 ¶ 1, Annex B.

[93] *See* Andrews, *supra* note 89, at A3.

[94] *See* Linda Gerstel et al., Joint Statement, On Occasion of the Final Plenary Meeting Concluding International Talks on the Preparation of the Foundation "Remembrance, Responsibility and the Future" (July 17, 2000) [hereinafter *Joint Statement*], *at* http://www.state.gov/www/regions/eur/holocaust/000717_joint_statement.html (last visited Nov. 25, 2001).

[95] BGBl. I S.1263.

[96] Law on the Creation of the Foundation § 20.

[97] The Bundestag declared legal peace on May 30, 2001. *See* Roger Cohen, *Last Chapter: Berlin to Pay Slave Workers Held by Nazis*, N.Y. TIMES, May 31, 2001, at A7. The Foundation began making payments on June 19, 2001. *See* Stephanie Flanders, *Payments Begin for Laborers Forced to Work for the Nazis*, N.Y. TIMES, June 20, 2001, at A10.

After Eizenstat and members of the Austrian government concluded a separate compensation agreement in October 2000, payments were similarly halted until pending cases were resolved. Once an Austrian lawyer agreed to dismiss a forced labor suit before a New York court, payments were scheduled to begin after July 19, 2001. *See Zwangsarbeit: Erste Zahlungen ab August*, DER STANDARD, July 19, 2001. Even after the payments began to flow, conflict persisted over the matter of the companies' fulfilling their obligations. *See* Christoph Irion, *Wirtschaft will Restbetrag in den Entschädigungsfonds einzahlen. Noch fehlen 580 Millionen Mark für frühere Zwangsarbeiter*, in BERLINER MORGENPOST, July 25, 2001.

[98] *See In re* Austrian and German Bank Holocaust Litigation, No. 98 CIV 3938 SWK, 2001 WL 228107 (S.D.N.Y. Mar. 8, 2001). These cases involved primarily property rather than labor claims. *See id.* at *2.

[99] *See* Mark John, *Nazi Slave Fund Suffers Legal Blow*, JERUSALEM POST, Jan. 26, 2001, at A9. Among Judge Kram's concerns were the possibility of a conflict of interest among plaintiffs' attorneys and the impact the settlement would have on absent claimants. *See id.* When asked about this issue, plaintiffs' counsel conceded the presence of a conflict

pressed astonishment. "This is devastating and frustrating to me and the people I represent," proclaimed Karl Brozik of the Jewish Claims Conference.[100]

Around this time, a new book entitled *IBM and the Holocaust* reached the shelves.[101] In it, journalist Edwin Black attempted to demonstrate that custom-built IBM technology enabled the Nazis to deport and exterminate with the efficiency and systemization for which they are so notorious.[102] In February, some of the same lawyers who had signed onto the Agreement the preceding July filed suit against IBM for its alleged complicity with the Nazi regime, relying chiefly on evidence presented in Black's book.[103]

Still awaiting a decision from Judge Kram, the Germans were incensed. "It is really annoying that a climate of uncertainty can grow out of new complaints that come precisely from where you don't expect it, namely those who signed the compensation agreement," Dr. Michael Jansen of the Foundation said to DeutschlandRadio.[104]

On March 8, 2001, Judge Kram issued her order denying the plaintiffs' motion for voluntary dismissal.[105] She set forth two reasons.

First, Judge Kram was concerned that dismissal would prejudice absent class members.[106] Normally, in a class action in which the class has yet to be certified, Judge Kram stated, "pre-certification dismissal does not legally bind absent class members," so the risk of prejudice is minimal.[107] This case, however, differed in three important respects:

> First, the [Agreement] provides that the Foundation is to be the exclusive forum for persons with claims against the German banking institutions that arise out of Nazi-era wrongs and atrocities. Second . . . the United States has agreed to issue a "Statement of Interest" urging dismissal of any and all new litigation against German entities for Nazi-era claims. Finally, the Foundation is not yet fully funded, and the parties submit that

---

of sorts, but argued that the conflict was typical of cases involving a capped fund. *See id.*

[100] *Id.*

[101] EDWIN BLACK, IBM AND THE HOLOCAUST (2001).

[102] *See generally id.*

[103] *See Suit Alleges IBM Aided Third Reich; Nazis Benefitted from Technology, Firm Says*, DALLAS MORNING NEWS, Feb. 12, 2001, at A3.

[104] *German Firms Not Getting Guarantees Against Slave Labor Lawsuits*, AGENCE FRANCE PRESSE, Feb. 28, 2001.

[105] *See In re* Austrian and German Bank Holocaust Litigation, No. 98 CIV 3938 SWK, 2001 WL 228107 (S.D.N.Y. Mar. 8, 2001). FED. R. CIV. P. 23(e) prohibits dismissal of a class action by consent of the parties alone; the court must approve the dismissal.

[106] *See In re* Austrian and German Bank Holocaust Litigation, No. 98 CIV 3938 SWK, 2001 WL 228107, at *5–*7 (S.D.N.Y. Mar. 8, 2001).

[107] *Id.* at *4 (citations and internal quotation marks omitted).

> such funding is contingent upon the Court dismissing the instant case.[108]

According to Judge Kram, the plaintiffs' motion could therefore be summed up as follows:

> the named plaintiffs ask the Court to voluntarily dismiss their claims, and thus subject all absent class members to the detrimental statement of interest and the other terms of the [Agreement], even though the absent class members' only source of compensation for their claims has yet to be fully funded.[109]

This set of circumstances, Judge Kram concluded, "constitutes unacceptable prejudice to the absent plaintiff class . . . and it would be unjust to divert their claims to a forum whose funding remains in question."[110]

The second reason cited by Judge Kram concerned the peculiar circumstances of one sub-class.[111] In a prior order, Judge Kram allowed a motion for voluntary dismissal by a sub-class with claims against Austrian banks.[112] One of the terms of the approved settlement provided for the assignment to the sub-class of any claims the Austrian banks may have had against German banks or other institutions that dominated them during World War II. The sub-class's original claims had thus been dismissed, leaving the sub-class no viable claim under the Foundation Law.[113] The class members' only remaining claims were now those of Austrian banks against German banks.

This limitation on relief for the sub-class troubled Judge Kram.[114] Allowing the voluntary motion to dismiss, Judge Kram reasoned, "would unduly prejudice the Assigned Claims sub-class," for it would permit plaintiffs's counsel to consent to a result—implementation of the Foundation Law as an exclusive remedy—that would not benefit the sub-class.[115]

Having set forth these two separate grounds, Judge Kram denied the motion, explicitly allowing for renewal of the motion "in the event full funding of the Foundation is accomplished, and the prejudice to the Assigned Claims sub-class is eliminated."[116]

---

[108] *Id.* at *5. Judge Kram also stated that the Statement of Interest was expected to be "a highly compelling and persuasive consideration to a court in favor of dismissal" according to the Special Master appointed to review the parties' accord. *Id.* at *6.

[109] *Id.* at *6.

[110] *Id.* at *7.

[111] *See id.* at *7–*8.

[112] *See In re* Austrian and German Bank Holocaust Litigation, 80 F. Supp. 2d 164, 180 (S.D.N.Y. 2000).

[113] *See In re* Austrian and German Bank Holocaust Litigation, No. 98 CIV 3938 SWK, 2001 WL 228107, at *7 (S.D.N.Y. Mar. 8, 2001).

[114] *See id.*

[115] *See id.* at *8.

[116] *See id.*

The Germans reacted with acrimony. Count Otto Lambsdorff, German industry's chief negotiator, called the decision "erroneous, bad, [and] unjustified."[117] Volker Beck, Green Party legal expert and a trustee of the Foundation, complained that the ruling "costs time that we don't have."[118] Executives at Deutsche Bank and DaimlerChrysler circulated angry letters to other German industry leaders arguing that no money should be paid while cases are still pending.[119] The United States Department of State also condemned Judge Kram's ruling, stating that it was "contrary to the recommendations of all parties having an interest in the case and will delay justice and payments to Holocaust victims, a significant number of whom are dying each month."[120]

Not everyone blamed the judge, however. As the *New York Times* reported, "[L]awyers representing Holocaust victims said ... that ... German industry had created its own problems [by refusing] to settle the claims as a conventional class-action lawsuit .... 'The reason we are in this position is that German industry and German parliamentarians insisted from the very first day on using the wrong legal mechanism,'[one lawyer said]."[121] Furthermore, some blamed the companies for failing to raise the full amount of their pledged contribution.[122]

In response to the latter criticism, a German industry spokesman announced on March 13, 2001, that the companies had raised their share of the Foundation money.[123] While six thousand companies had pledged to contribute by this time, it was the seventeen original companies that guaranteed the amount of the shortfall.[124] Still, the spokesman said, no payments would be made until legal peace had been achieved.[125]

---

[117] Edmund L. Andrews, *New Legal Disputes Put Holocaust Victim Payments in Doubt*, N.Y. TIMES, Mar. 9, 2001, at A3.

[118] *Federal Judge Declines to Dismiss Holocaust-Reparations Suit*, N.Y. TIMES, Mar. 8, 2001, at A8.

[119] Andrews, *supra* note 117, at A3.

[120] Press Statement, United States Dep't of State, spokesman Richard Boucher, United States Court Decision on Holocaust-Era Payments (Mar. 7, 2001).

[121] Andrews, *supra* note 117, at A3.

[122] *See German Industry Under Fire after U.S. Slave-Labor Ruling*, DESERET NEWS (SALT LAKE CITY, UT), Mar. 9, 2001 at A7.

[123] *See* Edmund L. Andrews, *Germans, Citing Suits, Say They're Holding War Slaves' Fund*, N.Y. TIMES, Mar. 15, 2001, at A11.

[124] *See* Herman, *supra* note 68 (reporting that the Foundation Initiative's Spokesman, Wolfgang Gibowski, confirmed that the founding companies of the Foundation Initiative would increase their contribution substantially in order to fulfill the Initiative's obligation to provide for its 50% share of the Foundation's DM 10 billion fund).

[125] *See German Industry at 2.5 Billion Euro Target for Slave Labor Fund*, AGENCE FRANCE PRESSE, Mar. 13, 2001. Another spokesman indicated that not every lawsuit had to be dismissed, but of the seventeen outstanding, all but the obscure claims had to go. *See* Stephen Graham, *Slave-Labor Compensation Flow Linked to Lawsuit Dismissals*, JERUSALEM POST, Mar. 15, 2001 at 1. United States Secretary of State Colin Powell sent a letter to the German Foreign Minister pointing out that fifty-five cases had been dismissed so far and that the United States was complying with its obligations under the Agreement. *See* Press Statement, U.S. Dep't of State Spokesman Richard Boucher, Contribution of German Industry to German Foundation, (Feb. 23, 2001), *at* http://www.state.gov/r/pa/prs/ps/2001/

On March 20, Judge Kram denied a motion to reconsider her ruling.[126] She told litigants that while she accepted that the Foundation was now fully funded, her concerns regarding the assigned claims sub-class were still outstanding.[127]

By this time, the parties in the suit against IBM had agreed to dismiss[128] and, as the *Frankfurter Allgemeine Zeitung* observed, just one "almost eighty-year old lady in New York" stood in the way of payments.[129] The parties not only appealed the ruling, but also filed a motion to disqualify the judge.[130]

Nearly two months later, with the Foundation fully funded and $2 million set aside for the assigned claims sub-class, Judge Kram finally capitulated.[131] In the May 11 order granting the motion to dismiss, Judge Kram indicated, however, that her order was based on the assumption that it would permit the German Bundestag to make a finding of "legal peace" and authorize the disbursement of Foundation funds "by the close of the present session of the Bundestag." [132] "If any of the assumptions on which this renewed motion is made are not realized or prove to be untrue," the judge wrote,

> plaintiffs have represented that they will file motions . . . to vacate the orders granting motions for voluntary dismissal. . . . Plaintiffs may also move . . . to vacate this Order in the event

---

index.cfm?docid=1254 (last visited Nov. 25, 2001).

[126] *See* Gail Appleson, *Judge Refuses to Dismiss Holocaust Suits*, JERUSALEM POST, Mar. 22, 2001, at 6.

[127] *See id.*

[128] *See* Press Statement, U.S. Dep't of State, spokesman Richard Boucher, IBM Lawsuit and Opening of Holocaust-era Corporate Archives (Mar. 29, 2001), *at* http://www.state.gov/r/pa/prs/ps/2001/index.cfm?docid=1783 (last visited Nov. 25, 2001).

[129] Jürgen Jeske, *Das Entschädigungsdebakel*, FRANKFURTER ALLGEMEINE ZEITUNG, Mar. 9, 2001, at 1 ("Nobody could have or wanted to imagine that, eight months later [after the Foundation Agreement of 17 July 2000, L.A./P.Z.] an almost eighty-year-old lady in New York with the will to fight would call into question the legal peace that had been established by a German-American Agreement." (transl. by Zumbansen)).

The international pressure to begin making payments cannot be overstated. *See, e.g.*, Roger Cohen, *Germany: Dispute on Holocaust Payments*, N.Y. TIMES, Apr. 7, 2001, at A4 (regarding Prime Minister Jerzy Buzek of Poland insisting in a meeting with Chancellor Gerhard Schröder that payments begin immediately).

[130] *See* Roger Cohen, *Germany: Holocaust Lawsuit*, N.Y. TIMES, Mar. 30, 2001, at A6.

The parties petitioned the Second Circuit for a writ of mandamus directing the district court to dismiss the case and also asking that the case be remanded to a different district court judge. *See* Duveen v. United States Dist. Court (In re Austrian & German Holocaust Litig.), 250 F.3d 156, 165 (2d Cir. 2001). The Second Circuit did not act on the petition until Judge Kram conditionally dismissed the case in May; the court found that at that point the petition met the criteria for either mandamus or an appeal, but ordered mandamus, remanding the case to Judge Kram. *See id.*

[131] *See* Jane Fritsch, *Judge Clears Obstacles to Pay Slaves of the Nazis*, N.Y. TIMES, May 11, 2001, at A15.

[132] Duveen, 250 F.3d at 161–62 (citing paragraph 4(b) of Judge Kram's order).

> the eligibility criteria of the German Foundation . . . are not revised [to protect the assigned claims sub-class].[133]

This conditional dismissal was not good enough for the parties, who sought relief before the Second Circuit on the grounds that Judge Kram had exceeded her authority in this portion of the opinion.[134] The Second Circuit agreed.[135] While cases and controversies are committed to the judiciary under Article III of the United States Constitution, the Second Circuit held, "[t]he conduct of foreign relations is committed largely to the Executive Branch."[136] The court pointed to the *political question doctrine*, which "restrains courts from reviewing an exercise of foreign policy judgment by the coordinate political branch to which authority to make that judgment has been 'constitutional[ly] commit[ted].'"[137]

The Second Circuit went on to observe that paragraph 4(b) of the district court order "seemingly requires the German legislature to make a finding of legal peace and to do so before its summer recess,"[138] and "paragraph 7 appears to indicate that if the German legislature failed to change the German law,"[139] it would be grounds for vacating the dismissal.[140] The district court does not have the authority, the Second Circuit found, "to require . . . the legislature of a foreign sovereign" to declare such things,[141] nor "is it the office of the [district] court . . . to decide what legislation should be enacted."[142] The Second Circuit remanded the case with instructions to Judge Kram to amend her order, eliminating paragraphs 4(b) and 7.

Following the Second Circuit decision, Chancellor Schröder predicted that payments would begin flowing soon, a Germany industry spokesman expressed confidence that adequate legal security had been realized, a spokeswoman for the Jewish Claims Conference welcomed the decision, and a Foundation trustee and member of the German parliament stated "[t]he time is ripe for payment."[143] On May 30, 2001, the Bundestag declared that legal peace had been attained, as was its charge

---

[133] *Id.* (citing paragraph 7 of Judge Kram's order).

[134] *Id.* at 162. Since the parties were of one mind before the Second Circuit, the district court defended its decision through counsel; representing the district court was attorney David Boies of Bush v. Gore, 531 U.S. 98 (2000), and Microsoft antitrust litigation fame.

[135] *See id.*

[136] *Id.* at 163.

[137] *Id.* at 164 (citations omitted). For analysis, *see infra* Part III.A.1.d.

[138] Duveen, 250 F.3d at 164.

[139] *Id.* at 165.

[140] *See id.*

[141] *Id.* at 164.

[142] *Id.* at 165.

[143] Steven Silber, *German Industry Okays Payments*, JERUSALEM POST, May 23, 2001, at 6.

under section 17 of the Law.[144] The Foundation began making payments the following month.[145]

D. Representational Shortcomings in the Negotiations and Limitations on Eligibility

The Law was designed to provide the exclusive avenue to recovery from the German government or German industry for injuries arising out of the Nazi labor program.[146] All claimants had to renounce future claims when they filed their applications for recovery.[147] Originally, most claims had to be filed within eight months of the Law's becoming effective (that is, by April 2001), but the deadline was eventually extended to the end of that year.[148] As a general matter, heirs are not eligible to recover, but the Law does permit recovery for heirs whose predecessors died after February 16, 1999, the date of the initial decision to establish the Foundation.[149]

While one might presume such restrictions to be the result of a negotiated compromise among affected parties, in fact representatives from key victim groups such as gay workers and disabled workers were absent from the negotiations.[150] An American lawyer named Barry Fisher joined discussions in the middle to speak to the interests of the Roma, but no one actually consented to the Agreement on their behalf.[151] Representa-

---

[144] Roger Cohen, *Last Chapter: Berlin to Pay Slave Workers Held by Nazis*, N.Y. TIMES, May 31, 2001 at A7.

[145] *See* Stephanie Flanders, *Payments Begin for Laborers Forced to Work for the Nazis*, N.Y. TIMES, June 20, 2001, at A10. In the time since the Foundation began making payments, some of the same attorneys have filed a new lawsuit claiming *inter alia* that the companies have failed to turn over the interest that accrued on the base amount (DM 10 billion) and that the claims assigned to the Austrian claim subclass are worthless. Telephone Interview with Deborah Sturman, Special Counsel, *In re* Holocaust Victims Assets Litigation (Nov. 29, 2001).

[146] Law on the Creation of the Foundation § 16(1), v. 2.8.2000 (BGBl. I S.1263) ("Payments from public funds, including social security, and from German business enterprises for injustice suffered under National Socialism as defined in Section 11 may be claimed only under the terms of this Law. Any further claims in connection with National Socialist injustices are excluded.") Note that this language may preclude recovery outside of the Law for abuses of the Nazi era other than slave and forced labor.

[147] *Id.* § 16(2).

[148] *Id.* § 14. This section provides an exception to the original deadline for non-Jewish claimants outside Central and Eastern Europe, applying instead a twelve-month deadline, presumably because these claimants will be most difficult to locate. *See Poles Start Receiving Payments for Slave Labor Under Nazis*, N.Y. TIMES, June 29, 2001, at A5.

[149] Law on the Creation of the Foundation § 13(1).

[150] *See* Stuart Eizenstat, Statement at Conclusion of Eighth Plenary Meeting of Slave and Forced Labor Negotiations, Washington, D.C. (Feb. 1, 2000), *at* http://www.state.gov/www/policy_remarks/2000/000201_eizenstat_labor.html. *See also Joint Statement*, *supra* note 94.

[151] *See id.* For an introduction to the Roma as well as an account of Roma experiences during and since their persecution under National Socialism, see Symposium, *supra* note 48, at 92–122. The Roma and Sinti were excluded from German compensation law from the beginning. German courts viewed prosecution of Roma and Sinti as legitimate law

tives from five Central and Eastern European countries attended the negotiations.[152] They found themselves in a weak bargaining position, however, and complained that the United States goaded them into accepting the terms in spite of their sense that the compensation amounts were too low.[153]

In addition, an untold number of potential plaintiffs who had not signed on with any of the lawsuits were necessarily unrepresented in the negotiations.[154] Perhaps they were thought to be represented in the manner of a class action.[155] In fact, many of the lawsuits that prompted the

---

enforcement activity well into the 1960s; as a result, they had no claim under the BEG. Cord Brügmann, *"Wiedergutmachung" und Zwangsarbeit. Juristische Anmerkungen zur Entschädigungsdebatte*, 16 DACHAUER HEFTE 177, 184 (2000).

[152] *See Joint Statement*, *supra* note 94. In addition to the United States, Germany, and Israel, the signatories include Belarus, the Czech Republic, Poland, Russia, and the Ukraine. *Id.*

[153] *See* Imre Karacs, *Germany's £3.2BN Bid to Close Book on Nazi Past Leaves only Rifts and Rancour in its Wake*, THE INDEPENDENT, July 18, 2000, at 3 ("Washington . . . said, 'If you don't like it we're going to sign anyway,' said Markiyan Demidov, head of the Ukraine's Union of Victims. 'They forced us to accept it. The amounts are a joke, an insult from Germany.'"); William Drozdiak, *Germany Sets Fund For Slaves Of Nazis; $5 Billion Will Go To Aging Survivors*, WASH. POST, July 18, 2000, at A17 ("Many survivors said they were distraught by the long delays and the relatively trivial compensation. 'It's not so much, when you consider the scale of our suffering, the bestial treatment we were subjected to and the wounds that stayed for life,' said Marian Nawrocki, head of the Association of Polish Victims of the Third Reich.").

[154] Of course, this depends on one's view of what constitutes representation. Eizenstat has equivocated on this question. Consider his comments during the Swiss banks negotiations: "I am not part of an advocacy group . . . . I was the chief enunciator of the [United States] government's position, which was that we wanted a settlement, we wanted to see justice—but we did not want to see sanctions imposed, or conditions put on the merger of Swiss banks." June D. Bell, *The Man in the Middle: Atlanta Son Stuart Eizenstat Balances Responsibilities of a Government Insider with Jewish Claims on His Loyalty*, ATLANTA JEWISH TIMES INTERNET EDITION at 8–9 (Feb. 19, 1999), *at* http://www.atljewishtimes.com/archives/1999//021999cs.htm (last visited Nov. 25, 2001).

Compare that to his comments at a briefing on the slave labor negotiations:

> QUESTION: Since the meeting today was primarily composed of country delegates and NGOs were generally not invited, my question is who is representing those refugees, displaced persons and other people who have emigrated from Central Europe and are currently U.S. citizens, or even those that may be found in other continents of the world? Is there a process by which those survivor groups can take part in these working groups, or will they have to file class actions in order to be part of the conversation?
>
> UNDER SECRETARY EIZENSTAT: I hope they won't have to file class actions. . . .
>
> I like to think, frankly, the U.S. government should represent the interests of U.S. citizens, regardless of their original nationality. And that is, indeed, who we are trying to represent.

Stuart Eizenstat, Briefing on establishing a process to achieve legal closure and make payments in connection with forced and slave labor and other outstanding claims under the Nazi regime, Washington, D.C., (May 12, 1999), *at* http://www.state.gov/www/policy_remarks/1999/990512eizenstat_slavelabor.html (last visited Nov. 25, 2001).

[155] *See* FED. R. CIV. P. 23.

negotiations were intended to be class actions.[156] Even assuming, however, that two classes of laborers from the United States, Israel, Central and Eastern Europe and elsewhere would have been certified,[157] the putative class members did not have an opportunity to "opt out" of the Law as they would have had in a traditional class action settlement.[158]

In a traditional class action settlement, "the judgment, whether favorable or not, will include all members who do not request exclusion."[159] As a general matter, "[i]f all of the requirements and prerequisites for a class action have been satisfied, the resulting decree will be binding on all class members whether they actually participated in the case or not."[160] Nevertheless, "an absent class member will not be bound if he can demonstrate" that one of the requirements has not been met.[161] If, for example, a class member was not provided with an opportunity to opt out of the action, he or she will not be bound.[162]

The requirements for a class action maintained under the Federal Rules of Civil Procedure, rule 23(b)(1)(B), however, differ from those for one maintained under subdivision (b)(3) in an important way: there is no opportunity to opt out.[163] The United States Supreme Court recently addressed the suitability of a case for treatment as a so-called "mandatory class action" in *Ortiz v. Fibreboard Corp.*[164]

Fibreboard concerned liability for personal injuries caused by exposure to asbestos.[165] In the 1970s the number of asbestos cases grew to an "elephantine mass," prompting the implicated companies to initiate negotiations in pursuit of a "global settlement" resulting in "total peace," i.e., no liability outside the purview of a single, mass settlement for $1.535 billion.[166] The district court certified a mandatory class and approved a global settlement and the Fifth Circuit affirmed.[167]

Class actions can be mandatory (i.e., without an opportunity to opt out) in cases in which "adjudications with respect to individual members of the class . . . would as a practical matter be dispositive of the interests of the other members not parties to the adjudications or substantially im-

---

[156] *See, e.g.*, Iwanowa v. Ford Motor Co., 67 F. Supp. 2d 424, 475 (D.N.J. 1999).

[157] *See* FED. R. CIV. P. 23(b).

[158] *See* FED. R. CIV. P. 23(c)(2).

[159] FED. R. CIV. P. 23(c)(2)(B).

[160] JACK H. FRIEDENTHAL ET AL., CIVIL PROCEDURE § 16.8 (2d ed. 1993).

[161] *Id.*

[162] *See id.* This is because "when a class action is maintained under subdivision (b)(3) . . . individual interest is respected. Thus the court is required to direct notice to the members of the class of the right of each member to be excluded from the class upon his request." FED. R. CIV. P. 23(c)(2) advisory committee's note (1966).

[163] *See* FED. R. CIV. P. 23(c)(2)–(3).

[164] 527 U.S. 815, 824 (1999).

[165] *Id.* at 821.

[166] *Id.* at 821–24.

[167] *Id.* at 828.

pair or impede their ability to protect their interests."[168] The Fifth Circuit found that the *Fibreboard* case presented just such a circumstance because the "limited fund" threatened the possibility of future recovery by any non-participating class member.[169]

The Supreme Court reversed, in part based on its rejection of the application of the "limited fund rationale."[170] In some cases, the Court conceded, "equity required absent parties to be represented, joinder being impractical, where individual claims to be satisfied from . . . one asset would . . . prejudice the rights of absent claimants against a fund inadequate to pay them all."[171] "'Classic' limited fund class actions," the Court wrote, involved things like "'trust assets, a bank account, insurance proceeds, [or] company assets in a liquidation sale.'"[172] In *Fibreboard*, however, the Supreme Court found that the record "failed to demonstrate that the fund was limited except by the agreement of the parties."[173]

In particular, the Court was troubled by the fact that "Fibreboard was allowed to retain virtually its entire net worth."[174] This "arrangement" the Court thought, "seems irreconcilable with . . . denying any opportunity for withdrawal of class members."[175]

After *Fibreboard*, it is difficult to imagine a lower court certifying a mandatory class under the Federal Rules of Civil Procedure 23(b)(1)(B) in a case in which the funds were limited only by agreement and the defendants included thriving German companies such as Daimler-Chrylser or Volkswagen. The Law, nonetheless, purports to bind all potential claimants in precisely this fashion. The terms of the Agreement,[176] the Law,[177] and the Statement of Interest[178] all indicate unequivocally that the Foundation Law henceforth provides the *exclusive* avenue to recovery. Thus, the Law is intended to preclude future claims, even though there is no settlement in the usual sense. Even the California plaintiffs are intended to be bound, despite the fact that some of their lawyers were denied a seat at the table.[179]

---

[168] FED. R. CIV. P. 23(b)(1)(B).

[169] *Fibreboard*, 527 U.S. at 829.

[170] *Id.* at 848–53.

[171] *Id.* at 836.

[172] *Id.* at 834 (citations omitted).

[173] *Id.* at 848.

[174] *Id.* at 859–60.

[175] *Id.* at 860. The Court also thought that certification of a mandatory class under such circumstances provided too great an incentive for global settlement and raised serious issues regarding the adequacy of representation, including possible conflicts of interest for class counsel. *Id.* at 852, 860 n.34.

[176] Agreement, *supra* note 2, at art. 1 ¶ 1.

[177] Law on the Creation of the Foundation § 16(1), v. 2.8.2000 (BGBl. I S.1263).

[178] *See* Agreement, *supra* note 2, at Annex B.

[179] Further, the Statement of Interest effectively voids California's law extending the statute of limitations for slave labor claims until 2010. *See infra* note 64 and accompanying text.

In short, individual survivors of the Nazi labor program brought their disputes before United States and German courts with the help of retained counsel, but their claims were settled by lawyers—sometimes not their own—in conjunction with representatives of some of the affected nations and ill-defined non-national constituencies. Following the Agreement, those whose attorneys and representatives were excluded from meaningful participation in negotiations, along with those who lacked representation entirely, are intended to be barred from pursuing recovery in any other forum.

The Law's eligibility criteria reflect similar deficiencies. The Law affords compensation only to victims who labored in industry or for public institutions.[180] Those who performed forced labor in ecclesiastical institutions, private households, or the agricultural sector have no claim under the Foundation Law, though the Law purports to preclude their lawsuits.[181]

The Foundation's money is being distributed by seven *partner organizations*.[182] These organizations are not agencies of any government but private entities that have contracted their services to the Foundation.[183] Most formed over the course of the 1990s (after the fall of the Iron Curtain) with the cooperation of the German government for the explicit purpose of distributing funds to survivors of Nazism.[184] The part-

---

[180] *See* Law on the Creation of the Foundation § 11(1).

[181] *See id.* Compensation for agricultural workers is left to the discretion of the partner organizations. *See id.*

[182] *Id.* §§ 9, 10. Section 9(2) assigns responsibility for specific regions to the seven partner organizations as follows:

> 1. Poland; 2. Ukraine and Moldova; 3. Russia, Latvia and Lithuania; 4. Belarus and Estonia; 5. The Czech Republic; 6. Non-Jewish claimants outside the states listed in 1–5 above; 7. Conference on Jewish Material Claims Against Germany for Jewish claimants living outside the states listed in 1-5 above.

Id. § 9(2).

[183] *See id.* § 10(1).

[184] The partner organizations were founded by interest groups and in cooperation with the German Government in the aftermath of German reunification and the dissolution of the Eastern bloc states. *See* WIEDERGUTMACHUNG UND KRIEGSFOLGENLIQUIDATION, *supra* note 43, at 189. Prior to the fall of the Soviet Union, many forced laborers might have been reluctant to bring claims. According to Ulrich Herbert, "a quite substantial number of Soviet prisoners of war and civilian workers were suspected by the Soviet authorities of having collaborated with the Germans . . . . It is surmised that there were severe repressive measures implemented against them after their repatriation." HERBERT, *supra* note 3, at 185.

The Conference on Jewish Material Claims Against Germany (the "Jewish Claims Conference") is an exception in that it formed not long after the conclusion of World War II. On October 26, 1951, delegates from twenty-two major Jewish national and international organizations convened in New York. *See* NANA SAGI, GERMAN REPARATIONS, A HISTORY OF THE NEGOTIATIONS 74–76 (1980). The purpose of the conference was to organize a unified Jewish stance that could secure reparations from Germany while ensuring that the State of Israel did not usurp Jewish representation during negotiations. *See id.* Jews from every continent except Asia and Antarctica met and declared that although Germany could never propitiate its moral debt through material reparations, justice de-

ner organizations have assumed responsibility for outreach to potential claimants in their respective constituency populations, determining the credibility of individual claims and disbursing funds.[185]

While section 11 of the Law sets forth eligibility standards for the partner organizations to follow, it provides scant guidance on how to evaluate the veracity of claims.[186] Similarly, the Law obligates partner organizations to establish appellate organs[187] yet provides no standards to govern appeals. Further, these organs must be "independent and subject to no outside instruction";[188] presumably, this language is intended to preclude judicial review of adverse decisions.[189]

---

manded the restoration of seized Jewish property and indemnification of victims of persecution. *Id.* at 75–76; CONFERENCE ON JEWISH MATERIAL CLAIMS AGAINST GERMANY, 1998 ANNUAL REPORT WITH 1999 HIGHLIGHTS at 25, 42 (1999); *see generally* SAGI, *supra*, at 76 ("No indemnity, however large, could make good the destruction of human life and cultural values. But every elementary principle of justice and human decency required that the German people should, at the least, restore seized Jewish property, indemnify the victims of persecution, their heirs and successors, and pay for the rehabilitation of the survivors."). The delegates established the Jewish Claims Conference as a representative body for future activity. *Id.* at 77.

[185] Law on the Creation of the Foundation §§ 9–10.

[186] *See id.* § 11(2) ("Eligibility shall be demonstrated by the applicant by submission of supporting material. The partner organization shall bring in relevant evidence. If no relevant evidence is available, the claimant's eligibility can be made credible in some other way.") Of course, the paucity of guidance available to decision-makers on how to evaluate claims could work to the benefit or detriment of an individual claimant, depending on the concrete circumstances of his or her claim.

[187] *Id.* § 19.

[188] *See id.*

[189] The Law's apparent prohibition against judicial review arguably is in tension with the German constitutional principle of "access to justice," according to which citizens have a right to bring claims before a judge. Article 10(2) of the German Basic Law (Germany's constitution) provides:

> Restrictions may only be ordered pursuant to a statute. Where a restriction serves the protection of the free democratic basic order or the existence or security of the Federation or a State [Land], the statute may stipulate that the person affected shall not be informed and that recourse to the courts shall be replaced by a review of the case by bodies and auxiliary bodies appointed by Parliament.

Grundgesetz [Basic Law] art. 10. (F.R.G.), *available in English at* http://www.uni-wuerzburg.de/law/gm00000_.html (last visited Nov. 25, 2001). *See also* BVerfGE 30 (1970), 1 (30–32) [first "G-10" decision] (holding that the installation of a parliamentary review council through a law providing for surveillance measures that touch upon the liberties protected under Art. 10 to control these surveillance measures is constitutional, as long as the individual reserves a right to bring the case before an administrative court); *see* also BVerfGE 100 (1999), 313 [second "G-10" decision] (holding that Art. 10 *GG* also protects the individual against the electronic use of the data collected in the process of legitimate surveillance); *see generally* Christoph J. M. Safferling, Zwangsarbeiterentschädigungsgesetz und Grundgesetz, 34 KRITISCHE JUSTIZ 208, 219 (2001) (arguing that access to justice and judicial review emanates from the principle of rule of law, as laid down in Art. 19 ¶ 4 art. 20 ¶ 1 of the German Basic Law); Ilse Staff, *Sicherheitsrisiko durch Gesetz. Anmerkung zum Urteil des Bundesverfassungsgerichts zum G-10 Gesetz*, 32 KRITISCHE JUSTIZ 586 (1999) (arguing that the FCC fails to draw distinctive lines between the competencies of surveillance agencies and those of the criminal prosecution). International agreements to which Germany is a party may also prohibit such competencies

In sum, the Foundation Law imposes strict temporal restrictions on recovery, denies compensation entirely to certain victim groups and most heirs, and purports to prevent claimants from contesting compensation denials before national courts. Moreover, these constraints did not receive the assent of those affected.

E. Rationales for the Foundation Law

In justifying to the Senate Foreign Relations Committee the United States government's decision to oppose legal claims by Holocaust victims, former United States Deputy Treasury Secretary and Special Representative for Holocaust Issues Stuart Eizenstat[190] argued that such opposition would benefit the victims themselves. "Conventional litigation," he contended, "would be a highly unsatisfactory solution for elderly slave and forced laborers and others injured by German companies during the War."[191] Eizenstat set forth three reasons for his claim:

> First, the success of litigation is problematic, given the variety of legal defenses available. Already, federal judges have dismissed two of the cases.
>
> Second, litigation would take years to reach fruition, with lengthy discovery, motions and appeals. Survivors average around 80 years of age and are passing away at a rate of some one percent a month. Thus, few survivors would benefit from litigation, even if it were successful [and]
>
> . . . .
>
> Third, any litigation would benefit only, at best, a small subset of surviving slave and forced workers, compared to the number who would benefit from the German Foundation Initiative . . . . [T]he Foundation Initiative will cover, under relaxed standards of proof, some one million workers, including those who worked for German companies now defunct or not subject

---

delegation without opportunity for judicial review. *See, e.g.*, International Covenant of Civil and Political Rights, Dec. 16, 1966, art. 14, 1966 U.S.T. Lexis 521 at 105–08; European Convention on Human Rights, Art. 6 (Nov. 4, 1950), *at* http://www.echr.coe.int/Convention/Convention%20countries%20link.htm (last visited Nov. 25, 2001).

[190] Though Eizenstat was appointed by President Clinton, he continued to perform the role of Holocaust Issues Representative at the invitation of President Bush. *See Federal Judge Declines to Dismiss Holocaust-Reparations Suit*, N.Y. TIMES, Mar. 8, 2001, at A8.

[191] *Eizenstat Testimony on Worldwide Holocaust Restitution Efforts: Hearing before the Senate Foreign Relations Comm.*, 106th Cong. (Apr. 5, 2000) (statement of Eizenstat, U.S. Deputy Sec'y and Special Rep. for Holocaust Issues).

> to U.S. jurisdiction including SS companies and companies owned by the German government.[192]

Eizenstat's arguments are compelling, suggesting that the Foundation, even with its capped amounts and limitations on eligibility and appeals, was the best deal that survivors could have struck. A closer look, however, reveals some shortcomings in this cluster of rationales.

First, litigation's alternatives have not been so speedy. Negotiations took a year and a half, followed by post-negotiation wrangling for another year.[193] At this writing, the funds still have not been entirely disbursed.[194] It is not self-evident that settlement and distribution confined to a judicial forum would have proceeded at a slower pace.

Furthermore, given the size of the sums available under the Foundation, it is not clear that the settlement was much more of a material victory for victims than an unsuccessful lawsuit would have been. On June 28, the day she received at most $2,200, Alicja Chyl of Poland told the Associated Press, "It's a piteous amount of money . . . . It's nothing for my work."[195] Aron Krell, a Polish-born survivor living in New York, lamented, "To me this is partial back pay—very little, very late . . . . Even if you said we were owed the minimum wage that was prevalent then in Germany, with a tiny rate of interest the amounts would have to be much, much larger than what we're getting."[196] It is not clear that a settlement is made better by arriving quickly or covering more people.

In any event, Eizenstat's claims prove too much; they could support any number of approaches to compensation, not just the Foundation Law. Germany and the United States could, for instance, have removed legal defenses and placed the litigation on an expedited schedule as has been used in other cases in which time was critical to satisfactory resolution.[197] Legislation could just as easily have been used to improve litigation's chances, as the legislature of California surely recognized when it extended the statute of limitations for insurance and slave labor claims.[198]

Our primary concern, however, is with Eizenstat's contention that the claimants were unlikely to prevail in court. This statement is important, not only because of its implications for the bargaining relationships

---

[192] *Id.*

[193] *See infra* Part II.

[194] Telephone Interview with Arie Bucheister, Director, Survivor's Assistance Program, Jewish Claims Conference (Nov. 28, 2001). Each partner organization is processing claims at its own pace; no specific deadline governs. *Id.*

[195] *Poles Start Receiving Payments for Slave Labor Under Nazis*, N.Y. TIMES, June 29, 2001, at A5.

[196] Stephanie Flanders, *Payments Begin for Laborers Forced to Work for the Nazis*, N.Y. TIMES, June 20, 2001, at A10.

[197] *See, e.g.*, Bush v. Gore, 531 U.S. 98 (2000). Alternatively, allowing heirs to recover in their predecessors' stead could have mitigated the urgency of relief.

[198] *See infra* notes 60–61 and accompanying text.

among participants in the negotiations, but also because it adds credence to German industry's refrain that it owes merely a moral and not a legal obligation.[199] This view is echoed in the preamble to the Foundation Law, which contemplates the "historic" responsibility of German industry and the "political and moral" responsibility of the entire country, but conspicuously neglects to assign legal responsibility to anyone.[200]

The notion that Germany and German companies owe every brand of responsibility except legal responsibility comes a little too easily and generally has been offered without any real discussion of the legal issues. In the next Part, we discuss the legal issues and their historical context in an effort to highlight the key political and moral judgments they implicate.

III. THE HISTORICAL CONTEXT

A. The Treaties

Since the 1950s, German courts have interpreted international treaties to preclude litigation by individual claimants. In this section, we dis-

---

[199] "The German companies have made it a great sense of pride that they're doing this as a moral, not a legal, gesture." Eizenstat press conference, Mar. 10, 2000.

We were unable to find any comments by Eizenstat regarding his own position on the matter of legal responsibility or on any of the legal questions raised by plaintiffs' claims, though he was asked about it on at least two occasions. On May 12, 1999 Eizenstat was asked whether the United States "recognize[s] any legal claims for forced labor, or is this really a moral affair?" He answered as follows:

> Well, there was a great deal of discussion, actually, by both sides on the issue of moral and legal. The plaintiffs' attorneys, as you would expect, talked a great deal about the legal nature of the claims. Interestingly, a number of survivor groups stressed the fact that it was the moral issues that were involved. The German companies and the German Government talked about a moral contribution, a voluntary contribution.
>
> I think that it's fine to talk in those terms. But quite frankly, to be practical, instead of trying to pigeon-hole whether it's a moral or legal issue, we have one basic goal in mind. That goal is to make the German initiative work. . . . I think, frankly, it's better to concentrate on that rather than to talk about, at this point, whether this is a moral or legal issue. It's obviously a combination of both and the two are almost inseparable.

At a White House briefing on December 15, 1999 (p. 102), Eizenstat was asked a related question: "Mr. Eizenstat, as you know, there were two lawsuit [sic] in New Jersey on slave labor claims that were dismissed. What did the dismissal of those suits—how did those dismissals impact the negotiations? And why didn't the United States intervene in those lawsuits with an amicus brief, with some kind of interpretation of the treaties that were being decided?" Eizenstat answered "I think I would have to leave it up to the lawyers to tell you what impact it had" and "we were not asked by those judges for our opinion, so we did not provide it."

[200] *See* Law on the Creation of the Foundation, Preamble, v. 2.8.2000 (BGBl. I S.1263).

cuss the relevant treaties and how they affect questions critical to the resolution of restitution claims.

German legislation dealing with restitution for stolen property and compensation for injustices committed during the Third Reich is generally referred to as *Wiedergutmachung* ("making good again").[201] Immediately after Germany's surrender and the creation of the Allied zones on German territory, an intensive "juridification"[202] in the field of restitution and compensation set in. Initially, American military law predominated,[203] but in the 1950s, the Bundestag passed several pieces of domestic legislation[204] in accordance with its obligations under the Transition Agreement of 1952 (Überleitungsvertrag).[205] The most significant of these, the Bundesentschädigungsgesetz (BEG),[206] had the declared purpose of providing compensation for injustices suffered under Nazi rule.

Several restrictions severely reduced the law's reach. First, the BEG imposed a territorial requirement. Only those victims who resided within

---

[201] For a critique of this term, see Karl Brozik, *Einmalig und voller Lücken: Entschädigung und Rückerstattung, in* TÄTER-OPFER-FOLGEN. DER HOLOCAUST IN GESCHICHTE UND GEGENWART 183, 183 (Heiner Lichtenstein & Otto R. Roberg eds., 1995); Sandro Blanke, *Der lange Weg zur Entschädigung von NS-Zwangsarbeitern*, 34 KRITISCHE JUSTIZ 195, 199 (2001). *See generally* Brügmann, *supra* note 151, at 180 n.9 ("The usage of the problematic term 'Wiedergutmachung' is often times justified by its having been used by Holocaust victims outside of Germany during the period of the Nazi rule. The term has been established."); HELGA & HERMANN FISCHER-HÜBNER, DIE KEHRSEITE DER "WIEDERGUTMACHUNG" 11 (1990).

[202] PAWLITA, *supra* note 43, at 3 (claiming that this juridification constituted a "state run, lawfully organized appropriation of history"). As Pawlita observes, the matter of compensation concerns the continuity or discontinuity of a state; thus, Germany's politics of compensation clearly had to be seen as a "parameter of its treatment of National Socialism." *Id.*

[203] *See* Hans-Dieter Kreikramp, *Zur Entstehung des Entschädigungsgesetzes der amerikanischen Besatzungszone, in* WIEDERGUTMACHUNG IN DER BUNDESREPUBLIK DEUTSCHLAND 61 (Ludolf Herbst & Constantin Goschler eds., 1989).

[204] *See* BUNDESRÜCKERSTATTUNGSGESETZ [Federal Restitution Law], v. 19.7.1957 (BGBl. I S.734); BEG, v. 26.6.1956 (BGBl. I S.550) (going into effect retroactively on October 1, 1953 and replacing BUNDESERGÄNZUNGSGESETZ [Federal Supplementary Law], v. 21.9.1953 (BGBl. I S.1387)). For a thorough and comprehensive account of how this legislation played out in administrative practice, see generally FISCHER-HÜBNER, *supra* note 201. *See also* Walter Schwarz, *Die Wiedergutmachung nationalsozialistischen Unrechts durch die Bundesrepublik Deutschland: Ein Überblick, in* WIEDERGUTMACHUNG IN DER BUNDESREPUBLIK DEUTSCHLAND 33 (Ludolf Herbst & Constantin Goschler eds., 1989); Ernst Féaux de la Croix, *Vom Unrecht zur Entschädigung: der Weg des Entschädigungs-rechts, in* DIE WIEDERGUTMACHUNG NATIONALSOZIALISTISCHEN UNRECHTS DURCH DIE BUNDESREPUBLIK DEUTSCHLAND, VOL. III, 1, 1–118 (Bundesjustizminister der Finanzen [German Finance Ministry] & Walter Schwarz eds., 1985).

[205] *See* Brügmann, *supra* note 151, at 181.

[206] BUNDESENTSCHÄDIGUNGSGESETZ [Federal Compensation Law] (BEG), v. 26.6.1956 (BGBl. I S.559), at §§ 1, 43(2). The BEG recognized claims only for personal and asset injury due to persecution for political, racial, religious, and ethnic reasons, and excluded all potential claimants living in states to which Germany had no diplomatic ties, including (at the beginning of the Cold War) Eastern European countries. *Id.* at § 238a. Those countries, however, had provided a majority of the slave and forced laborers. *See* Sandro Blanke, *Der lange Weg zur Entschädigung von NS-Zwangsarbeitern*, 34 KRITISCHE JUSTIZ 195, 199–200 (2001).

German territory, who were stateless, or who belonged to the group of exiled, deported or "expelled" persons, and who could establish the required territorial connection to the German Reich within the 1937 parameters, were eligible.[207] Second, the BEG did not recognize slave or forced labor per se as grounds for compensation.[208] As most of the slave and forced laborers had come from Eastern Europe,[209] and many survivors returned to these countries after their liberation, most former laborers had no claim under the BEG.[210] Third, the BEG-Schlussgesetz of 1965[211] imposed an expiration date, barring claims made under the BEG after 1969.

At the time the original BEG legislation was enacted, Germany expressed fear to the international community that individual compensation for each of the Third Reich's victims would be financially impossible.[212] As a result, collective compensation emerged as the favored approach, and Germany became nearly immune to individual suits.[213] Even in the immediate aftermath of World War II, therefore, litigation for individual plaintiffs held little promise.

In addition to the Transition Agreement of 1952, the German government entered into a number of other post–World War II treaties.[214] The earliest treaties primarily concerned security matters.[215] Parties to subsequent agreements, however, quickly realized that plans to integrate postwar Germany with the anti-Soviet West would have significant repercus-

---

[207] BEG §§ 4, 167.

[208] *Id.* at § 43(2). The BEG recognized claims for deprivation of freedom, injury, or death that might arise in connection with forced labor, but forced labor itself did not entitle a person to recovery. *Id.* Compensation also was provided for general persecution because of race, religion, political opinion, furthermore for professional damage ("*Berufsschaden*"), i.e., the termination of one's career. *Id. See also* Walter Schwarz, *Zur Einführung: Das Recht der Wiedergutmachung und seine Geschichte*, 26 JURISTISCHE SCHULUNG 433, 437–39 (1986); Wiedergutmachung und Kriegsfolgenliquidation 82–88 (HERMANN-JOSEF BRODESSER ET AL. EDS. 2000).

[209] Herbert, *supra* note 52, at 221–29; Benz, *supra* note 52, at 4; PAWLITA, *supra* note 43, at 34–38, 48–52.

[210] BEG §§ 4, 167.

[211] Zweites Gesetz zur Änderung des Bundesentschädigungsgesetzes (BEG-Schlussgesetz), v. 14.9.1965 (BGBl. I S.1315) (amending the BEG).

[212] *See* BGH, RzW 14 (1963), 525 (526); BGH, NJW, 35 (1973), 1549 (1550).

[213] Blanke, *supra* note 206, at 200; Brügmann, *supra* note 151, at 183.

[214] *See* PAWLITA, *supra* note 43, at 110, 146. These treaties followed the Yalta and Potsdam Agreements between Victorious Nations of February and July/August 1945 and precede the London Debt Agreement of 1953. *See* Abkommen über deutsche Auslandsschulden, v. 27.2.1953 (BGBl. II S.333) (available in English, French, and German). Among the most prominent are the Paris Reparation Treaty of 1946 and the Peace Treaties of 1947 with Italy, Romania, Hungary, Bulgaria and Finland. *See* Hess, Völker- und zivilrechtliche Beurteilung der Entschädigung für Zwangsarbeit vor dem Hintergrund neuerer Entscheidungen deutscher Gerichte, in: Entschädigung für NS-Zwangsarbeit 65, 81 (1998); Wiedergutmachung und Kriegsfolgenliquidation. Geschichte—Regelungen—Zahlungen (Hermann-Josef Brodesser, et al. eds. 2000), *supra* note 43, at 63-64.

[215] PAWLITA, *supra* note 43, at 110, 146.

sions on prospects for reparations.[216] If Germany were burdened with onerous reparation obligations, the likelihood of a speedy economic and political recovery would be slim.[217]

As a result, Germany's obligations following World War II differed substantially from its obligations under the Treaty of Versailles after World War I.[218] The post–World War II scheme reflected more cautious reparation politics. Under the German-Israeli Treaty of 1952 (known as the Luxembourg Agreement),[219] Germany agreed to pay DM 3 billion to the State of Israel,[220] while under an agreement with the Jewish Claims Conference concluded the same day, Germany paid DM 450 million.[221] A series of bi- and multilateral international agreements signed between 1954 and 1964 set free nearly DM 900 million in pension and one-time compensation payments to former victims of Nazi persecution now residing in eleven Western signatory states.[222] German law professor Burkhard Hess recently noted that "both the compensation paid under the BEG as well as under global compensation agreements have pushed aside holding the individual companies liable."[223]

The preference for global over individual compensation prevented most of the Third Reich's victims from recovering. Non-Germans, who formed a majority of those persecuted by the Nazis, were considered residents of former enemy states and were largely excluded from compensation.[224] Global compensation also eliminated opportunities for individualized compensation. Courts regularly dismissed individual lawsuits by referring to existing compensation legislation; indeed, courts held that compensation legislation and international treaties precluded the claims

---

[216] *Id.* at 71–87.

[217] *Id.*

[218] It would not be unfair to view this result as a windfall for Germany, given the traditional use of the reparations instrument by formerly warring nations. Ignaz Seidl-Hohenveldern, *Reparations*, *in* 4 ENCYCLOPEDIA OF PUBLIC INT'L LAW 178, 178–80 (2000).

[219] Abkommen zwischen der Bundesrepublik und Israel, Sept. 10, 1952, art. 1–5, v. 21.3.1953 (BGBl. II S.35).

[220] *See* Karl Brozik, *Einmalig und voller Lücken: Entschädigung und Rückerstattung*, *in* TÄTER-OPFER-FOLGEN. DER HOLOCAUST IN GESCHICHTE UND GEGENWART 183, 187 (Heiner Lichtenstein & Otto R. Roberg eds., 1995). For an analysis of the differing views of the Treaty, see generally PAWLITA, *supra* note 43, at 269–84.

[221] *See id.* at 269, 284–89.

[222] Brügmann, *supra* note 151, at 182; ULRICH HERBERT, ARBEIT, VOLKSTUM, WELTANSCHAUUNG: UBER FREMDE UND DEUTSCHE IM 20 JAHRHUNDERT 177 (1995) (citing BT-Drs. 6287/30 (parliamentary report of Oct. 31 1986)); *Bundestagsdrucksache* (Parliament report of 31 October 1986) 10/6287, p. 30/50. These agreements drew both praise and disapproval, inside and outside Germany. *See, e.g.*, Michael Wolffsohn, *Globalentschädigung für Israel und die Juden? Adenauer und die Opposition in der Bundesregierung*, *in* WIEDERGUTMACHUNG IN DER BUNDESREPUBLIK DEUTSCHLAND, *supra* note 203, at 161; Norbert Frei, *Die deutsche Wiedergutmachungspolitik gegenüber Israel im Urteil der öffentlichen Meinung der USA*, *in* WIEDERGUTMACHUNG IN DER BUNDESREPUBLIK DEUTSCHLAND, *supra* note 203, at 215.

[223] Burkhard Hess, *Entschädigung für NS-Zwangsarbeit vor US-amerikanischen und deutschen Zivilgerichten*, 44 DIE AKTIENGESELLSCHAFT 145, 151 (1999).

[224] *See* Herbert, *supra* note 222, at 178–81.

even of plaintiffs who failed to meet the eligibility criteria of the compensation laws.[225]

The compensation regime enacted by the German parliament shortly after the war thus amounted to domestic statutory law compensating a small minority of victims and a series of agreements with foreign governments and the Jewish Claims Conference. The most critical diplomatic agreement for our purposes is the London Agreement on Germany's External Debts ("London Debt Agreement" or "LDA"). Concluded February 27, 1953,[226] the LDA was contemporaneous with Germany's compensation negotiations with Israel and the Jewish Claims Conference.[227] According to its preamble, the main purpose of the LDA was to enable Germany to establish normal economic relations with other nations.[228] The LDA contemplated a series of payments by the German government to the signatory states (the United States, Great Britain, France and the Soviet Union) but postponed reparation claims by these states until Germany regained its economic footing:

> Consideration of claims *arising out of the Second World War* by countries which were at war with or were occupied by Germany during that war, and by nationals of such countries, against the Reich, or *agencies of the Reich* . . . shall be deferred until the final settlement of the problem of reparations.[229]

German courts interpreted this provision to postpone various plaintiffs' claims until conclusion of "a final reparation treaty."[230] The German courts therefore rejected claims made before a final reparation settlement because they were "derzeit unbegründet" ("as yet unfounded").[231] Consequently, the length of time between the events giving rise to the claims and the adjudication of those claims grew.

To this day, no formal peace treaty has ever been concluded between Germany and the former Allies. The so-called "2+4 Treaty" of 1990[232]

---

[225] *See, e.g.*, BGH, RzW, 12 (1963), 525; BGH, NJW, 35 (1973), 1549 (1550).

[226] Abkommen über deutsche Auslandsschulden, v. 27.2.1953 (BGBl. II S.333) (available in English, French, and German).

[227] *See* HERBERT, *supra* note 222, at 163.

[228] *See* BGH, NJW, 35 (1973), 1549 (1551–52).

[229] Abkommen über deutsche Auslandsschulden, v. 27.2.1953 (BGBl. II S.333), at Article 5(2) (emphasis added).

[230] Entscheidungen des Bundesgerichtshofes in Zivilsachen (BZHZ) 18 (1955), 22 (22, 27, 32). *See also* BGH, Versicherungsrecht (VersR), 22 (1964), 637; BGH, RzW, 14 (1963), 525; BGH, NJW, 35 (1973), 1549 (1550).

[231] *Id.* at 1552 (noting the ill effects a reparation agreement would have on Germany's recovery, especially if compensation claims for forced labor figured among these reparations). *See, e.g.*, BGH, Monatsschrift des Deutschen Rechts (MDR), 17 (1963), 492; OLG Stuttgart, RzW, 15 (1964), 425; LG Frankfurt, NJW, 13 (1960), 1575 (1576); FERENCZ, *supra* note 8, at 132.

[232] The *2* refers to West Germany and East Germany, and the *4* refers to the United States, Great Britain, France and the Soviet Union. Vertrag über die abschließende Rege-

("2+4"), however, has come to serve the basic functions of a peace treaty.[233] This document, which was signed at the time of German reunification, restored total sovereignty to the newly unified Germany.[234] It makes no mention of reparations arising out of World War II. The question therefore arose whether 2+4 fulfills the vision of the London Debt Agreement. As § 12 of the preamble of 2+4 states, the signatory parties understand the treaty to conclude World War II-related matters with Germany;[235] 2+4 therefore seems to preclude further reparation treaties among the signatory states, notwithstanding the LDA's being entirely mute on the issue of reparations.

Since it eliminated the obstacle posed by the LDA, one might have expected that 2+4 would have spurred a wave of lawsuits. Instead, the significance of 2+4 and the possibilities it presented took time to sink in—too much time for many potential plaintiffs in Eastern Europe.[236] Uncertainty remained as to whether the LDA's concept of state-to-state reparations would still block individual forced labor claims.

At the same time, two lawsuits were launched against Germany before the Bremen and Bonn *Landgerichte* (Regional Courts) in an effort to break through the dual barriers erected by the restrictive BEG and the casting of compensation claims as reparations under the LDA.[237] After 2+4, it became clear that there would be no final reparation agreement. It seemed the only way to reopen avenues for compensation was to challenge Germany's post-war claim that compensation outside the purview of the BEG threatened to put Germany into insolvency.

In 1993, the Bonn Court, dealing with claims by plaintiffs who allegedly were excluded from recovery under national compensation laws like the BEG, presented to the *Bundesverfassungsgericht* (FCC) [federal constitutional court] the question of whether section 1 of the 1957 law

---

lung in Bezug auf Deutschland [Treaty on the Final Settlement with Respect to Germany], v. 12.9.1990 (BGBl. II S.1317) (including treaty text in English as well as German).

[233] *See* Dietrich Rauschning, *Beendigung der Nachkriegszeit mit dem Vertrag über die abschließende Regelung in bezug auf Deutschland, in* DEUTSCHES VERWALTUNGSBLATT (DVBl.) 105 (1990), 1275 (1279); Dieter Blumenwitz, *Der Vertrag vom 12.9.1990 über die abschließende Regelung in bezug auf Deutschland*, NJW, 48 (1990), 3041; PAWLITA, *supra* note 43, at 468–71.

[234] Vertrag über die abschließende Regelung in bezug auf Deutschland, v. 12.9.1990 (BGBl. II S.1317).

[235] *See* Rauschning, *supra* note 233, at 1279.

[236] *See, e.g.*, OLG Stuttgart, NJW, 53 (2000), 2680 (rejecting a claim for compensation in tort with reference to the three-year statute of limitations in German tort law, since the statute was activated on the day the 2+4 treaty went into force, on March 15, 1991). WIEDERGUTMACHUNG UND KRIEGSFOLGENLIQUIDATION. GESCHICHTE REGELUNGEN ZAHLUNGEN (Hermann-Josef Brodesser/Bernd Josef Fehn/Tilo Franosch/Wilfried Wirth eds. 2000), *supra* note 43, oppose this view, arguing that as 2+4 was mute on the matter of reparations but, instead, was aimed at erecting a durable peace agreement for Europe, *id.* at 185–86. Nevertheless, Brodesser et al. conclude that the political developments and the German payments since 1945 had rendered another reparation agreement superfluous.

[237] *See* Burkhard Heß, *Entschädigung für NS-Zwangsarbeit vor deutschen und US-amerikanischen Gerichten*, 44 DIE AKTIENGESELLSCHAFT 145, 151 (1999).

Allgemeines Kriegsfolgengesetz (AKG) [General Law on the Consequences of the War][238] was constitutional. Section 1 of the AKG had been interpreted to preclude any claims against Germany not expressly recognized in that statute. This lawsuit presented the question of whether the political justifications offered by the German government at the law's enactment forty years earlier still held true, since there remained no doubt about Germany's successful economic recovery.

The FCC ruled that neither economic recovery nor German reunification necessitated a review of decisions made in the context of immediate post-war consolidation.[239] While the FCC acknowledged the possibility of revising the AKG's provisions, it considered this prerogative to reside with the German legislature. The FCC upheld section 1 of the AKG, explicitly cognizant of the historical and political circumstances that led to the AKG's curtailment of compensation claims.[240]

The Court added in *dictum* that no rule of international law impeded the legislature if it wished to recognize individual compensation claims, even if such claims were grounded in facts falling under the rubric of reparations (i.e., state-to-state compensation).[241] This portion of the FCC's decision was received by plaintiffs' representatives as a triumph,[242] but only for a brief period.

In 1997, following the FCC's decision, the Bonn Court awarded victory to a Polish woman, who (like twenty-one other plaintiffs involved in the suit) had taken on Israeli citizenship after the war. The court noted that § 8 para. 1 of the BEG prevented recovery by the other plaintiffs who had received prior compensation, yet held that the Polish woman had established a tort claim against Germany under § 839 of the German Civil Code (BGB).[243] The court found forced labor to be a clear case of violation of personal freedom and damage to health. Section 839 of the *BGB* and Art 34 *Grundgesetz* (German Basic Law), created a cause of action in tort against the State authority—here the SS at Auschwitz—for coercing the plaintiff into the forced labor.

Germany appealed the decision, and the Appeals Court reversed, finding the lone appellee's claim also barred by § 8 para. 1 of the BEG.[244] The Appeals Court held that the BEG's residency requirement (§ 4) ex-

---

[238] Gesetz zur allgemeinen Regelung durch den Krieg und den Zusammenbruch des Deutschen Reiches entstandener Schäden (Allgemeines Kriegsfolgengesetz), v. 5.11.1957 (BGBl. I S.1747).

[239] Entscheidungen des Bundesverfassungsgerichts ("BVerfGE"), 94 (1996), 315 (325).

[240] *Id.*

[241] BVerfGE, 94, 315 (331–34).

[242] *See id.* at 330–34. *But see* Hugo J. Hahn, *supra* note 1 (failing even to mention the decision when referring to the governing law in this field).

[243] LG Bonn, 16 STREIT 101 (1998), *overruled by* OLG Koln, 52 NJW 1555 (1999), *appeal docketed* BGH, IX ZR 439/98 (JURIS); *cf.* Herbert Küpper, Die neuere Rechtsprechung in Sachen NS-Zwangsarbeit, in 31 KRITISCHE JUSTIZ 246 (1998).

[244] *See* OLG Köln, 52 NJW 1555, 1555 (1999).

cluded the appellee from recovery, *and* that § 8 para 1 made her ineligible for any *other* compensation. According to the Appeals Court's decision, compensation for Nazi-era persecution under the BEG (which does not specifically recognize forced labor as a claim per se) was the only avenue to compensation. Having failed to meet the residency requirement, the woman was effectively blocked from claiming any compensation from Germany, even outside of the BEG. The appeal before the Federal Court of Justice (*Bundesgerichtshof*) is, at this writing, still pending.

In sum, Nazi-era laborers asked German courts to examine whether existing law blocked avenues for individual compensation outside the BEG. Courts consistently held that it did. Though the BEG regime denied compensation on the basis of slave or forced labor, limited the number of eligible claimants, and precluded all claims against Germany outside the BEG, German courts held that the choice of a particular compensation regime was a matter for the legislature which courts could not review. Thus the German case law interpreting international treaties and national legislation reflects the same choice as the Foundation Law: it diverts claims from the courtroom to the legislature. Moreover, courts not only declined to assign legal responsibility, but also obscured any traces of the post-war political judgments that guided their decisions or those of the German legislature.

B. Problems Raised by the Treaties

1. Statute of Limitations

Cases brought before German courts in the past few years have confronted a charge of untimeliness according to the applicable statute of limitations.[245] These cases, like their predecessors, reveal a highly fragmented understanding of the applicability of this total bar. For instance, a lawsuit brought in 1959 by a German national against his former "employer" was rejected on the ground that his claim had expired.[246] This case established a two-year statute of limitations by tying the compensation claim to the civil law provisions dealing with pay for work, notwithstanding the fact that the plaintiff had not consented to his employment.[247] Since this decision, the precise legal category appropriate for a

---

[245] *See, e.g.*, LG Hamburg, NJW, 52 (1999), 2825; LG Berlin, NJW, 53 (2000), 1958; OLG Hamm, NJW, 53 (2000), 3577.

[246] *See* Entscheidungen des Bundesgerichtshofes in Zivilsachen (BZHZ), 48 (1967), 126 (126). Plaintiff asserted that he had worked under gruesome conditions for more than two years for the defendant without pay; the defendant had instead paid the monthly sum of 132 Reichsmark to the Nazis. The Court denied the necessity of employment-at-will for an assessment of plaintiff's claim for wages, ruling that plaintiff had worked under a *de facto* contract. The Court rejected both the unjust enrichment claim for wages and the tort claim on the ground that they had expired.

[247] *See* Heß, *supra* note 237, at 151 (noting simply that the "Bartl" case had been re-

claim of enslavement has changed from case to case, and is reflected in the varying limitation periods that German courts have applied.[248]

In the 1990s, German courts continued to apply different statutes of limitations. While some courts began with the assumption that there is a general two-year statute in the context of employment, the *Bundesarbeitsgericht*, the highest employment court in Germany, refused to recognize any employment relationship in a forced labor case and referred the case back to the civil courts.[249] The *Arbeitsgericht* Koblenz, a regional employment court, denied the existence of an employment contract between a Polish former forced laborer and the company for which the labor was performed, referring the case to the civil courts. The court also noted, however, "that civil law is made to react neither to the millions of cases of forced labor in this century nor to the takings, deportations and other mass crimes against life and health."[250]

In a 1999 case before the *Landgericht* Hamburg, the court held that claims brought this late regardless of the envisioned foundation solution had expired in 1979, i.e., thirty years after the enactment of the German Basic Law.[251] Evaluating the merits of the lawsuit to determine whether the court would be obliged to provide legal aid, the court raced through several alternative bases for the suit before dismissing it on statute of limitations grounds. Even if the generally applicable thirty-year statute of limitations[252] governed the claims, the plaintiff had failed to file suit before this thirty-year period expired. The clock had begun ticking the last time the plaintiff performed forced labor, or, at the very latest, at the time the *Grundgesetz* (German Basic Law) was enacted. The court concluded by observing: "[a]fter all, judicial aid cannot be granted, although for moral reasons which in the judicial aid procedure, however, cannot be of penetrating importance, claimant—if he has not been indemnified under the BEG—should at least, if he meets the requirements, receive compensation as a Polish forced laborer under the to-be-created Foundation Act,

---

jected because of the two-year statute of limitations).

[248] *See, e.g.*, BGH, BGHZ 48, 125 (1967) (2 years, § 196 Abs. 1 Nr. 9 *BGB*); LG Hamburg, NJW, 38 (1999), 2825 (30 years, § 195 *BGB*); Landesarbeitsgericht (LAG) [labor court of appeals] Nürnberg. Arbeit und Recht (AuR), (1999), 405; LAG München, AuR, (1999), 449; LAG Hessen, AuR, (1999), 450; Arbeitsgericht (ArbG) Hannover, AuR, (1999), 451; Bundesarbeitsgericht (BAG) [supreme labor court], NJW, 19 (2000), 1438 (all four cases: denying employment character and referring to trial court); BAG, Arbeit und Arbeitsrecht (AuA), 48 (2000), 228 (followed by critical commentary by Achim Seifert); OLG Stuttgart, NJW, 36 (2000), 2680.

[249] BAG, NJW, 19 (2000), 1438. *See also* BAG, AuA, 48 (2000), 228 (followed by Achim Seifert's critical commentary, which emphasizes that the Court neglects a number of legal considerations which would have produced the opposite result).

[250] ArbG Koblenz, NJW, 38 (1999), 2838 (2839) (trans. by Zumbansen).

[251] LG Hamburg, NJW, 38 (1999), 2825, 2825 (trans. by Zumbansen).

[252] *See* § 852 Abs.1 2. Alt. *BGB* (§ 852 para. 1 alt. 2 of the German Civil Code).

a Fund that is designed—as is commonly known—to preclude individual suits."[253]

Former slave and forced laborers were thereby left with no path to compensation. Prior to 2+4, German courts read the language of the London Debt Agreement (LDA) (deferring claims "arising out of the Second World War" until a "final settlement of the problem of reparations") to defer slave and forced labor claims. They regarded claims arising out of slave and forced labor as claims for reparations, or state-to-state payments, rather than as matters suited to private compensation for individuals.[254]

Though this might have been the case only until a final reparation agreement was concluded, after 2+4, claims were rejected for new reasons. The political forces that guided the LDA and the German compensation laws of the 1950s cast a long shadow on the prospects for litigating individual claims, even after 2+4. In more recent cases, claims for compensation have been caught between the post–World War II political impulse to avoid repeating the purported mistakes of Versailles, and the global settlement (the Foundation), which provides compensation for those who meet the criteria but excludes all individual claims outside its purview.

A conspicuously brief decision handed down by the FCC on April 25, 2001 illustrates and reinforces this trap.[255] The plaintiff asserted that section 16 of the Foundation Law deprived him (and other former slave and forced laborers) of the right to sue, in violation of Article 14 of the German Basic Law.[256] His argument could succeed, the court reasoned, only if the Foundation Law deprived him of an otherwise valid legal claim.[257] The FCC held that the complainant did not have a compensation claim outside of the one provided by the Foundation Law. Section 16 of the Foundation Law, therefore, did not collide with Article 14 of the German Basic Law. The FCC, certainly aware of the bluntness of such an assertion, based its decision on the existence of "not a single case" in which a forced labor claim had prevailed.[258] Notwithstanding the intricacies of the treaties, the national compensation legislation, and the case

---

[253] LG Hamburg, NJW, 52 (1999), 2825 (2825).

[254] *See infra* Part III.A.

[255] BverfG, NJW, 54 (2001), 2159.

[256] *See* Safferling, *supra* note 189, at 211–17.

[257] *Id.*

[258] BverfG, NJW, 54 (2001), 2159. The FCC apparently overlooked the Wollheim case. After winning in a lower court, Wollheim settled the suit for the sum of DM 30 million, DM 26.25 million of which were distributed to Jewish victims, while the remaining DM 3.75 million went to non-Jewish forced laborers of Polish descent. *See* Wolfgang Benz, *Der Wollheim Prozeß: Entschädigung für Zwangsarbeit in Auschwitz*, *in* ZWISCHEN HITLER UND ADENAUER: STUDIEN ZUR DEUTSCHEN NACHKRIEGSGESELLSCHAFT 128, 153 (1991).

law, the FCC declined to confront the complexities of the forced labor issue, readily telling a story of failure.

In the German context, this decision probably marked the end of "litigating the Holocaust"[259]—at least for the slave and forced labor claims. The decision by Germany's highest court does not leave much room to pursue avenues of relief other than the one offered by the Foundation.[260]

Given these decisions, we do not contend that additional litigation would have resulted in victories for the survivors of the Nazi labor program. We merely illustrate the political forces at work throughout the legal reckoning with slave and forced labor during the Third Reich. From the earliest legal decisions, one can discern the impact of the Versailles Treaty, which laid a heavy financial and political burden on Germany after World War I; the beginning of the Cold War; and the Allied Treaty politics with regard to international reparations and the resulting exclusion of individual claims.[261] Moreover, courts concluded in the 1960s and 70s that corporate actors were *agencies of the Reich*[262] within the meaning of the LDA,[263] thereby extending the LDA's deferment of claims to protect the companies on the basis of these historical-political influences.

Years later, judges deliberating on forced labor claims paid no attention to the circumstances under which post-war and Cold War judges had handled initial cases arising out of the National Socialist period. When the FCC allotted a mere two pages for a case of immense historical dimensions, declining to consider the complexities of the LDA and 2+4 and focusing exclusively on the outcomes of the earlier claims, it failed to account for the real animus behind the decisions upon which it relied. The court's decision represents the ultimate denial of victims' legal claims and a culmination of years of decisions adverse to the victims of Nazism. Their losses were counted in the record of stare decisis, but the political context, and therefore the real significance of the losses, was erased.

---

[259] *See* Brown, *supra* note 15, at 553.

[260] BVerfG, NJW, 30 (2001), 2159.

[261] The fact that many pre-war judges remained in office after 1945 may also have contributed. *See, e.g.*, INGO MÜLLER, FURCHTBARE JURISTEN: DIE UNBEWÄLTIGTE VERGANGENHEIT UNSERER JUSTIZ 203 (1989); Bernhard Diestelkamp, *Rechts–und verfassungsgeschichtliche Probleme zur Frühgeschichte der Bundesrepublik Deutschland*, 21 JURISTISCHE SCHULUNG 492 (1981); Günter Frankenberg & Franz J. Müller, *Juristische Vergangenheitsbewältigung—Der Volksgerichtshof vorm BGH*, 17 KRITISCHE JUSTIZ 145 (1983); Joachim Perels, *Die Restauration der Rechtslehre nach 1945*, 18 KRITISCHE JUSTIZ 359 (1984); Rolf Lamprecht, *Lesarten für Rechtsbeugung*, 47 NEUE JURISTISCHE WOCHENSCHRIFT 562 (1994).

[262] We mean that private entities engaged in the employment of forced and slave laborers were considered to be "agencies of the Reich" in the meaning of Art. 5(2) of the LDA and thereby immunized against individual claims until the signing of a peace treaty with Germany. For analysis, see *infra* Part III.B.2.

[263] *See* BGH, RzW, 14 (1963), 525; BGH, NJW, (1973), 1549.

2. Justiciability of Private Claims

Like the German courts, two federal district courts in New Jersey found that the LDA and 2+4 barred plaintiffs' claims. After a lengthy examination of the multiple post-war treaties from 1945 through 2+4 in 1991,[264] the *Degussa* court found that "individual claims for forced labor against the German government and private industry were subsumed by national governments in the treaties which concluded the war with Germany."[265] The court reached this conclusion notwithstanding plaintiffs' efforts to distinguish their private claims from claims for reparations arising out of war.[266] The *Iwanowa* court agreed.[267] In the view of both courts, plaintiffs' claims raised political questions that were not fit for judicial resolution.

To reach these conclusions, the New Jersey courts had to answer (explicitly or implicitly), two related questions inherent in their application of the LDA. First, did the plaintiffs' claims *arise out of* World War II? Second, were the companies acting *as agencies of the Reich* when they utilized slave and forced labor? An affirmative answer to both would mean that the LDA and 2+4 combined first to defer and then to preclude altogether plaintiffs' claims. We discuss each below ((a) and (b), respectively).

a. Arising out of War (Reparations vs. Compensation)

As German historian Ulrich Herbert, a leading expert in the field of slave and forced labor, declared in the summer of 2000 (just after Germany and German industry agreed to fund the Foundation with DM 10 billion), "[o]ne ought to keep in mind that forced labor had—for the longest time and until very recently—not been considered a [Nazi-era] crime as such. Instead forced labor had been regarded as an inevitable war related effort to confront war time labor shortages."[268]

In the early 1940s, there was already widespread belief that Germany would be called to account for the actions of its government and industries, and that a far-reaching restitution and compensation program was likely.[269] Nevertheless, it was still far from evident how it would be

---

[264] Burger-Fischer v. Degussa AG, 65 F. Supp. 2d 248, 262–72 (D.N.J. 1999).

[265] *Id.* at 281.

[266] *See infra* Part III.B.3 for a discussion of justiciability and the political question doctrine.

[267] Iwanowa v. Ford Motor Co., 67 F. Supp. 2d 424, 460 (D.N.J. 1999).

[268] Ulrich Herbert, *"Ein Element der Verunsicherung, der Irritation, des Erschreckens," Der Umgang mit der NS-Vergangenheit und die Entschädigung von Zwangsarbeitern*, 45 BLÄTTER FÜR DEUTSCHE UND INTERNATIONALE POLITIK 555, 558 (2000) (trans. by Zumbansen).

[269] FISCHER-HÜBNER, *supra* note 201, at 11. *See also* PAWLITA, *supra* note 43, at 160–91 (discussing various proposals by the Allied governments and by Jewish authors).

executed. Well into the post-war years, the distinctions between reparations and compensation in the case of Nazi Germany were only vaguely contemplated, eventually resulting in a dynamic that strongly favored state-to-state reparation over private compensation models.[270]

Furthermore, lawyers practicing compensation law did so in relative isolation. As post-war Germany focused on economic recovery and putting its past behind it, such lawyers found themselves marginalized by their colleagues.[271] The political climate in Germany in the 1950s was one of recovery, rebuilding and integration into the west—*not* one of turning back, regret or open discussion of the past.[272]

Even today, one would have difficulty finding a lawyer who can comfortably distinguish between reparations and compensation. Only a handful of lawyers have ever practiced in this field.[273] The situation for plaintiffs was further complicated by the heterogeneity of the claimants and claims. Individuals, organizations and states lodged competing claims for asset restitution, individual and collective compensation, and

---

[270] *See* HERBERT, *supra* note 222, at 160–61 (describing Germany's intent to subsume compensation payments for slave and forced laborers under a state-to-state reparation scheme); *see also* Hess, *Völker- und zivilrechtliche Beurteilung*, *supra* note 214, 86–89 (describing that while Germany succeeded in reaching a reparations moratorium with the London Debt Agreement, it was still obliged to provide compensation by national legislation); *see also supra* Part III.A. (regarding the scope of this legislation, namely the restrictive BEG, effectively excluding many potential recipients of compensation).

This was the case notwithstanding the fact that Germany enacted a substantial restitution scheme. *See* SPOERER, *supra* note 80, at 242 (2001) (regarding DM 85 billion distributed under the BEG); PETER REICHEL, VERGANGENHEITSBEWÄLTIGUNG IN DEUTSCHLAND: DIE AUSEINANDERSETZUNG MIT DER NS—DIKTATUR VON 1945 BIS HEUTE 74 (2001) (regarding an overall sum of DM 160 billion distributed under the *signum* of *Wiedergutmachung*).

[271] *See, e.g.*, PAWLITA, *supra* note 43, at 5. *See also* Schwarz, *supra* note 32, at 114. In this farewell note, Editor-in-Chief Schwarz noted that the weekly journal was neither read nor regarded as a promising publication by mainstream lawyers in post-war Germany. On occasion of the journal's last issue in 1981, Schwarz bitterly recalled that only four tenured professors chose to publish pieces in the journal and that, of the scant feedback the journal received, most was harsh criticism of the journal's publication policy, which was alleged to cast Germany's compensation politics in bad light. *Id.* at 115.

[272] *See* Schwarz, *supra* note 32; Blanke, *supra* note 208, at 201. Germany's treatment of its Nazi past has been the subject of intense research. *See, e.g.*, EUGEN KOGON, DIE UNVOLLENDETE ERNEUERUNG: DEUTSCHLAND IM KRÄFTEFELD 1945–1963 (1964); NORBERT FREI, VERGANGENHEITSPOLITIK: DIE ANFÄNGE DER BUNDESREPUBLIK UND DIE NS-VERGANGENHEIT (1996); LUTZ NIETHAMMER, DEUTSCHLAND DANACH: POST-FASCHISTISCHE GESELLSCHAFT UND NATIONALES GEDÄCHTNIS 53, 53–58 (1999) (discussing the forms of "de-nazification" in the four occupation zones); ULRICH HERBERT & OLAF GROEHLER, ZWEIERLEI BEWÄLTIGUNG: VIER BEITRÄGE ÜBER DEN UMGANG MIT DER NS-VERGANGENHEIT IN DEN BEIDEN DEUTSCHEN STAATEN (1992); HELMUT DUBIEL, NIEMAND IST FREI VON DER GESCHICHTE: DIE NATIONALSOZIALISTISCHE VERGANGENHEIT IN DEN DEBATTEN DES DEUTSCHEN BUNDESTAGES (1999); ALEIDA ASSMANN & UTE FREVERT, GESCHICHTSVERGESSENHEIT, GESCHICHTSVERSESSENHEIT: VOM UMGANG MIT DEUTSCHEN VERGANGENHEITEN NACH 1945 (1999).

[273] *See* Walter Schwarz, *Zur Einführung: Das Recht der Wiedergutmachung und seine Geschichte*, 26 JURISTISCHE SCHULUNG 433 (1986), at 433; *see also*, Abschied, *supra* note 32, at 115.

reparations.[274] As the German political scientist Peter Reichel has observed in his recent book on "Vergangenheitsbewältigung" ("the coming to terms with the past"), individual victims faced a politically difficult struggle in competing against these rival claims.[275] Acknowledgment of the post-war decision to avoid repeating the perceived mistakes of Versailles relieved Germany and German industry of any obligation to respond to individual compensation claims; the claims were simply subsumed under the concept of state-to-state reparations.[276] By grounding their decisions in this conceptualization of the claims, as well as in early interpretations of the treaties reflecting this conceptualization, German courts effectively conceded the historical contingency of their reasoning.

This concession, however, has done little for plaintiffs. Given the choice between accepting compensation along with an acknowledgement of "political, moral, and historical" responsibility under the Foundation Law or receiving nothing at all because "no case is known where plaintiffs succeeded,"[277] the choice for plaintiffs seems obvious.[278] If German courts had decided not to treat the use of slave and forced labor as war-related, plaintiffs might have presented claims for individual relief outside the context of state-to-state reparations.

b. Agencies of the Reich

As a related matter, companies have argued that, when they employed forced laborers, they were not autonomous employers in a private labor market, but were instead "agencies of the Reich."[279] The claims against them, they therefore argue, should meet with the demise contemplated under the LDA.

The language of the LDA appears to contemplate broad incorporation of private actors into a zone of impunity by postponing all claims

---

[274] REICHEL, *supra* note 270, at 76.

[275] *Id.* at 76.

[276] *See* HERBERT, *supra* note 222, at 160.

[277] Such was the formula reiterated by the *Bundesverfassungsgericht* in its dismissal of April 25, 2001. BVerfG, NJW, 54 (2001), 2159.

[278] *See* DEBORAH STONE, POLICY PARADOX: THE ART OF POLITICAL DECISION MAKING 242–56 (1997). Stone advises readers to "be on the lookout for Hobson's choices. Whenever you are presented with an either/or choice, you should be tipped off to a trap. You can disengage it by . . . expanding the range of consequences you bring into the analysis." *Id.* at 256.

[279] The famous (yet unpublished) Wollheim case before the Frankfurt Regional court on November 3, 1951, was one of the very rare instances in which this line of defense was not accepted by the judge. For an intriguing background account, see Benz, *supra* note 258, at 128–54; *see also* Wolfgang Benz, *In Sachen Wollheim gegen I. G. Farben*, 9 DACHAUER HEFTE 142–47 (1993); *cf.* Otto Küster, *Das Minimum der Menschlichkeit: Plädoyer*, 9 DACHAUER HEFTE 156–74 (1993) (one of Wollheim's lawyers providing the argument to the court). In contrast, the defendant companies in *Iwanowa* argued that they were *not* de facto state actors for purposes of the Alien Tort Claims Act. Iwanowa v. Ford Motor Co., 67 F. Supp. 2d 424, 443 (D.N.J. 1999).

until the time of a final peace treaty. Certainly when a country is at war, the private sector is heavily implicated. No doubt this was the case for employers (and employees) in wartime manufacturing and production. One *could* adopt a position that would render all private actors *agents* as long as their actions were connected to the war. To the extent, however, that the applicability of the term *agent* relies on the companies having been *coerced* by the German wartime authorities to employ forced labor, the historical evidence is, at best, disputed.[280] In either event, if the companies were agents of the Reich, according to the conventional German construction of the LDA, the claims against the companies would be deferred. Then, with the signing of 2+4 nearly forty years later, that deferment would become permanent.

3. The Political Question Doctrine

Similar issues arose before the Second Circuit when it invoked the *political question doctrine* and determined that Judge Kram had exceeded her authority by imposing conditions in her order of voluntary dismissal.[281] The court explained, "the political-question doctrine restrains courts from reviewing an exercise of foreign policy judgment by the coordinate political branch to which authority to make that judgment has been constitutional[ly] commit[ted]."[282] Judge Kram's decision to condition dismissal on the Bundestag's prompt declaration of legal peace as well as its amending the Law to protect the assigned claims sub-class, therefore, caused the Second Circuit "considerable difficulty."[283]

The political question doctrine was famously expounded by Justice Brennan in *Baker v. Carr*.[284] In that case, which concerned a Tennessee reapportionment statute, the Court set forth six criteria, any of which would render a question nonjusticiable and all of which share the aim of

---

[280] *See, e.g.*, FERENCZ, *supra* note 8, at 33 (reproducing a translated version of a 1943 letter from the chairman of the board of I. G. Farben to Chief of Police Heinrich Himmler requesting expansion of a factory "as was done at Auschwitz"). Thousands of similar documents were submitted in the criminal trials of I. G. Farben's officers at Nuremberg to establish that the company bore "[r]esponsibility for taking the initiative in the unlawful employment." *Id.* at 34.

The decision to define the term "agent" broadly for this purpose is, of course, a moral one, not compelled by the legal materials. *See* Brügmann, *supra* note 151, at 186.

[281] *See* Duveen v. United States Dist. Court (In re Austrian & German Holocaust Litig.), 250 F.3d 156 (2d Cir. 2001).

[282] *Id.* at 164 (citations omitted).

[283] *Id.* Recall, however, that the claims arose in the first instance in a judicial forum. The parties became embroiled in international diplomacy and legislation by a foreign sovereign in order to resolve pending litigation. Once the terms of resolution were brought before the court for approval, it seems strange to deem them suddenly nonjusticiable. Judge Kram merely issued an order regarding the case before her. She did not purport to order the Bundestag to act; she spoke to the adequacy of resolution of the claims pending in her court.

[284] 369 U.S. 186, 208–37 (1962).

ensuring that federal courts avoid upsetting the balance of powers by intruding on questions best left to the political branches.[285]

This vexatious doctrine[286] also turned up to rationalize dismissal in the New Jersey cases, *Degussa* and *Iwanowa*.[287] In deciding to dismiss claims against Degussa and Siemens, Judge Debevoise observed: "Every human instinct yearns to remediate in some way the immeasurable wrongs inflicted upon so many millions of people by Nazi Germany so many years ago, wrongs in which corporate Germany unquestionably participated . . . ."[288] Nevertheless, Judge Debevoise applied the six-prong

---

[285] *See id.* at 217. Justice Brennan explained:

> Prominent on the surface of any case held to involve a political question is found a textually demonstrable constitutional commitment of the issue to a coordinate political department; or a lack of judicially discoverable and manageable standards for resolving it; or the impossibility of deciding without an initial policy determination of a kind clearly for nonjudicial discretion; or the impossibility of a court's undertaking independent resolution without expressing lack of the respect due coordinate branches of government; or an unusual need for unquestioning adherence to a political decision already made; or the potentiality of embarrassment from multifarious pronouncements by various departments on one question.

Id. The Court found that none of these six criteria justified a finding of nonjusticiability with regard to Tennessee's statute. *See id.*

In what appears to be dicta, the *Baker* Court discussed several lines of cases that implicate the political question doctrine, including cases involving "[f]oreign relations" and "[d]ates of duration of hostilities." *Id.* at 211, 213. In these sections, Justice Brennan stated, for example, that "it is error to suppose that every case or controversy which touches foreign relations lies beyond judicial cognizance," but also that "recognition of foreign governments . . . strongly defies judicial treatment." *Id.* at 211–12. Furthermore, "the war power includes the power 'to remedy the evils which have arisen from its rise and progress,'" though "deference rests on reason, not habit." *Id.* at 213 (citations omitted). A question that does "not seriously implicate considerations of finality," i.e., whether hostilities have ceased, may be considered. *Id.*

[286] Whether the political question doctrine constitutes prudent judicial deference to the political branches for the sake of our constitutional framework or an anti-democratic abdication of judicial responsibility, especially as applied to foreign affairs, has been the subject of extensive scholarly debate. *See, e.g.*, David J. Bederman, *Deference or Deception: Treaty Rights as Political Questions*, 70 U. COLO. L. REV. 1439 (1999) (arguing that there is reason to be concerned about excessive judicial deference in the area of foreign affairs); Michael J. Glennon, *Foreign Affairs and the Political Question Doctrine*, 83 AM. J. INT'L L. 814 (1989) (arguing that the political question doctrine is an abdication of judicial responsibility); Louis Henkin, *Is There a "Political Question" Doctrine?*, 85 YALE L.J. 597 (1976) (criticizing the political question doctrine as unnecessary); Anne-Marie Slaughter (Burley), *Are Foreign Affairs Different?*, 106 HARV. L. REV. 1980 (1993) (reviewing THOMAS FRANCK, POLITICAL QUESTIONS/JUDICIAL ANSWERS: DOES THE RULE OF LAW APPLY TO FOREIGN AFFAIRS? (1992) and refuting Franck's insistence that the political question doctrine is as much an abdication of judicial responsibility in the context of foreign affairs as it is in other contexts).

[287] We will focus in our discussion on the application of the political question doctrine in *Burger-Fischer v. Degussa* AG, 65 F. Supp. 2d 248, 272–85 (D.N.J. 1999). *Accord* Iwanowa v. Ford Motor Co., 67 F. Supp. 2d 424, 483–89 (D.N.J. 1999).

[288] *Degussa*, 65 F. Supp. 2d at 285.

Baker analysis and concluded that the court lacked "the power to engage in such remediation."[289]

Like the German courts, Judge Debevoise concluded that the plaintiffs' claims were not justiciable because they arose out of war:

> Beginning with the Versailles Treaty concluding World War I, the term "reparations" has been deemed to refer to "all the loss and damage to which . . . Governments and their nationals have been subjected as a consequence of the war imposed upon them . . . ." Specifically, "reparations" include "[d]amages caused to civilians by being forced by Germany or her allies to labour without just remuneration."[290]

Claims for reparations, according to conventional international law, are reserved to governments; individuals have no claims to reparations.[291]

The plaintiffs in *Degussa* maintained, however, that they were not seeking reparations, but merely compensation for a private wrong. Plaintiffs' expert, Dr. Christian Wolf, "argue[d] that in the context of World War II forced labor claims are different from forced labor claims arising out of World War I and should not be included in reparation claims." Dr. Wolf distinguished between typical "'acts of war'" and "'unprecedented acts of extermination . . . based on racial motives.'" The relationship between them, Dr. Wolf asserted, was a "'mere temporal coincidence.'" Explaining that "reparations" are associated with war, Dr. Wolf concluded that the claims before the court concerned something different—namely "'racial and ideological persecution under the Nazi regime.'"[292]

Judge Debevoise was not convinced. He concluded, "[T]he forced labor program was primarily an act of war designed to enable the Third Reich to pursue its war of conquest."[293] According to the court, the fact that many of the laborers were victims of racial persecution imprisoned in concentration camps "did not change the nature of the program as primarily a war related effort, subject to reparations as negotiated by the victorious nations."[294]

---

[289] *Id.*

[290] *Id.* at 275.

[291] *Id.* at 274 (citing RESTATEMENT (THIRD) OF FOREIGN RELATIONS, § 902 cmt. (h)(I) (1987) and LOUIS HENKIN, FOREIGN AFFAIRS AND THE CONSTITUTION 262 (1972)).

[292] *Id.* at 275–76. *Cf.* RAUL HILBERG, THE DESTRUCTION OF THE EUROPEAN JEWS 643–62 (1961); HERBERT, *supra* note 3, at 158 (identifying a conflict between the dual goals of exterminating Jews on the one hand and waging war on the other; limited resources made it difficult to pursue both aims simultaneously).

[293] *Degussa*, 65 F. Supp. 2d at 276.

[294] *Id.* Judge Debevoise appears to have found Dr. Wolf's argument more compelling as to the concentration camp inmates from Germany than as to those laborers who were forcibly deported from conquered countries into the Reich. *See id.*

While Dr. Wolf's argument did not persuade the court, his analysis casts doubt on the determinacy of the *arising out of war* formula that would purport to classify plaintiffs' claims as either justiciable or non-justiciable on political question grounds. On the one hand, the forced labor program was no doubt conceived in part to further the Nazi war of aggression by sustaining armament and other production while much of the German labor force was otherwise occupied with military obligations.[295]

On the other hand, the Nazis embraced an extensive and complicated ideology of racial hierarchy.[296] This hierarchy structured the forced labor program, determining who was subject to it, for which jobs they would be eligible, and how they would be treated.[297] As the Reich's racial policy changed, so changed the labor program. After the Final Solution was declared at the Wannsee Conference in 1942, decisively adopting a policy of extermination (rather than deportation) of Jews,[298] the labor program instituted "annihilation by labor" for Jews as a consequence, harmonizing the dual goals of wartime production and extermination.[299]

Moreover, the conventional historical explanation for the decision to implement the forced labor program—that the Nazis were responding to a labor shortage as a result of German men having military obligations—is itself questionable. Nazi ideology precluded German women from entering the industrial work force; they were expected instead to stay at home and tend to the reproduction of the race.[300] Among non-Germans,

---

[295] *See* FERENCZ, *supra* note 8, at 17; HERBERT, *supra* note 3, at 132.

[296] *See id.* at 162–63. Herbert explains how the nuances of the Nazi racial ideology translated into living conditions for various workers:

> The workers from western countries were in a worse condition than native Germans, but in a far better position than workers from the East when it came to food rations, the interior furnishings and physical state of living quarters . . . . A hierarchy arose on this basis . . . French civilian workers ranked above all others . . . after these came workers from southern European countries. . . .Those workers were followed on a lower rung by laborers from Czechoslovakia . . . and beneath them were the Poles. At the bottom of the heap were workers from the Soviet Union—along, since the summer of 1943, with Italian military internees . . . . The concentration camp inmates . . . formed their own special category far beneath the bottom rung of this racist ladder, since their exploitation . . . was not tempered by any sort of treatment ordered [toward] productivity on the job.

Id. at 162–63.

[297] *See id.* at 163–64.

[298] *See* FERENCZ, *supra* note 8, at 16; *see also* HERBERT *supra* note 3, at 177.

[299] *See* FERENCZ, *supra* note 8, at 17–18 (regarding "extermination by labor"); HERBERT, *supra* note 3, at 177. Herbert argues that "[t]he thesis that the policy toward the Jews had a central primary aim—namely, their exploitation as forced laborers—is untenable. Instead it is more accurate to contend that the policy of the Final Solution was implemented beneath the camouflaging cover of forced labor deployment." *Id.* Later, Herbert explains more generally, "[r]acism was not some sort of mistaken belief serving to mask and cloak the true interests of the regime, which were in essence economic. Rather, racism was the very lodestar of the system, its unwavering fixed point." *Id.* at 179.

[300] *See* HERBERT, *supra* note 3, at 132. *But see* MARTIN JAY, THE DIALECTICAL IMAGINATION: A HISTORY OF THE FRANKFURT SCHOOL AND THE INSTITUTE OF SOCIAL RE-

the percentage of women expected to work in a particular national group bore an inverse relationship to that group's position in the Nazis' racial hierarchy; *i.e.*, the lower a group ranked, the greater the percentage of its female population could be found in the forced labor pool.[301]

We co-taught a seminar on the Nazi labor program and its legal fallout to German law students at the University of Frankfurt in December, 2000.[302] The students uniformly appeared to have accepted the truism that the program arose in response to a labor shortage. When we pointed out to them that World War II was an occasion for women in the United States to join the industrial work force at a higher rate than ever before due to the military obligations imposed on men,[303] they responded blankly at first, but then with the dawning realization that ideology played a role in the truism that there had been a labor shortage. Half of the German labor pool went untapped. The alleged labor shortage was due not to any fixed condition of war, but rather to constraints imposed by a racist and sexist ideology governing domestic life.

As it turns out, therefore, the Nazi labor program can be viewed as either war-related or as a matter of racial ideology. Judge Debevoise's decision was neither compelled by historical evidence nor conclusively rebutted by it. War as well as racist ideological motives propelled the program and gave rise to plaintiffs' injuries, so neither the legal conclusion that plaintiffs were seeking reparations nor the contrary conclusion that they were seeking compensation for a private wrong was required.

In adopting uncritically the explanation of wartime necessity rather than focusing on the ideological source of the wrongs suffered by the plaintiffs, Judge Debevoise already began deliberating on the merits of the plaintiffs' claims, not merely their justiciability. He addressed, perhaps unwittingly, some of the most difficult questions associated with the Holocaust: questions about causation, motive and agency.

To decide the *merits* of the claims, he and a jury would have had to sort through such thorny questions as whether the companies used slave labor because they were compelled by the Nazis to maintain production quotas; whether the companies took advantage of cheap labor primarily for profit; whether the companies adopted Nazi ideology and whether

---

SEARCH, 1923–1950, 247 (1973) describing work conducted by researchers associated with the Frankfurt School and reported in the famous study *The Authoritarian Personality* suggesting that the mother in the authoritarian family did not provide the "warmth and protectiveness [that] had once served as a buffer against the arbitrary harshness of the patriarchal world," and that the Nazis "undermin[ed] the family, despite their propaganda to the contrary."); *see also* THEODOR W. ADORNO ET AL., THE AUTHORITARIAN PERSONALITY (1950).

[301] HERBERT, *supra* note 3, at 153.

[302] *See German Law Journal Co-Editor, Dr. Peer Zumbansen, Leads Trans-Atlantic Seminar on Nazi Slave Labor Compensation at University of Frankfurt*, 2 GERMAN L.J. (Jan. 15 2001) (describing seminar), *at* http://www.germanlawjournal.com/past_ issues.php?id=48.

[303] *See, e.g.*, STEPHANIE COONTZ, THE WAY WE NEVER WERE: AMERICAN FAMILIES AND THE NOSTALGIA TRAP 159–60 (1992).

abuse of the slaves by the companies' German employees established the companies' adoption of the ideology; and so on. Even before reaching the merits, on a motion to dismiss on grounds of nonjusticiability, Judge Debevoise addressed these same sorts of questions by attributing the wrongs to war.

Our contention is not that the decision to categorize the plaintiffs' claims as reparations was an incorrect legal conclusion, but rather that it was not legally required. Legal ambiguity is important here because it gives rise to the question of how Judge Debevoise reached his decision. While ostensibly predicated on the necessities of law, his decision was instead driven by what might just as fairly be characterized as moral and political judgments, his own as well as those embodied in decades of German cases grounded in post-war politics.[304] Even at this early procedural phase, Judge Debevoise touched on vital questions in a matter of civil liability: What motivated the defendants to engage in the disputed conduct? How much choice did they have? How much responsibility can they fairly be assigned? Whether the plaintiffs' claims were justiciable, therefore, was not merely a threshold question, but really went to the ultimate question of culpability, whether described in legal or moral terms.

[304] To ensure the consistency of his decision, Judge Debevoise reviewed some of the same German case law we have reviewed. *See* Burger-Fischer v. Degussa AG, 65 F. Supp. 2d 248, 279–81 (D.N.J. 1999).

C. Substantive Defenses Historically Used by German Defendants

It is the historian's task to uncover "the reality of forced labor" many years after the actual occurrence.[305] While this task has met many obstacles—both psychological[306] and evidentiary[307]—the recent opening up of companies' archives[308] promises further revelations. Improved access to historical documentation will no doubt prove integral to the assessment of state and private responsibility.[309]

---

[305] For historical research, see, for example, ULRICH HERBERT, FREMDARBEITER, *supra* note 52; SPOERER, *supra* note 80; REINHOLD BILLSTEIN ET AL., WORKING FOR THE ENEMY. FORD, GENERAL MOTORS AND FORCED LABOR IN GERMANY DURING THE SECOND WORLD WAR (2000); GERALD D. FELDMAN, UNTERNEHMENSGESCHICHTE DES DRITTEN REICHS UND VERANTWORTUNG DER HISTORIKER: RAUBGOLD UND VERSICHERUNGEN, ARISIERUNG UND ZWANGSARBEIT (1999).

[306] For an analysis of both reluctance and outright anger about confronting the Nazi past among Germans in the aftermath of WW II, see generally ALEIDA ASSMANN & UTE FREVERT, GESCHICHTSVERGESSENHEIT, GESCHICHTSVERSESSENHEIT: VOM UMGANG MIT DEUTSCHEN VERGANGENHEITEN NACH 1945 (1999); Ulrich Herbert, *Zweierlei Bewältigung*, *in* ZWEIERLEI BEWÄLTIGUNG: VIER BEITRÄGE ÜBER DEN UMGANG MIT DER NATIONALSOZIALISTISCHEN VERGANGENHEIT IN DEN BEIDEN DEUTSCHEN STAATEN 9–10 (Ulrich Herbert & Olaf Groehler eds., 1992) (declaring that the denazification "experiment" had failed for a number of reasons, among which figured the sheer immensity of such an undertaking and the fact that the processes started in the Nuremberg aftermath against war criminals were stopped due to a changed political climate and the circumstance of a dawning cold war). *See also* NORBERT FREI, VERGANGENHEITSPOLITIK: DIE ANFÄNGE DER BUNDESREPUBLIK UND DIE NS-VERGANGENHEIT 25 (1996) (describing a "politics of the past" pursued by conservatives who, shortly after the German Bundestag's taking session, already aimed at the conclusion of the de-nazification politics); HELMUT DUBIEL, NIEMAND IST FREI VON DER GESCHICHTE: DIE NATIONALSOZIALISTISCHE HERRSCHAFT IN DEN DEBATTEN DES DEUTSCHEN BUNDESTAGES 39–40 (1999) (describing how early references to the Nazi crimes were disassociated from individual actors and, instead, attributed to the "criminal system" as such); Christopher R. Browning, *German Memory, Judicial Interrogation and Historical Reconstruction: Writing Perpetrator History from Postwar Testimony*, *in* PROBING THE LIMITS OF REPRESENTATION: NAZISM AND THE "FINAL SOLUTION" 25 (Saul Friedlander ed., 1992) (addressing the intricacies of writing about the events from an historiographic perspective).

[307] Compare, for example, the varying indications as the total numbers of forced laborers under Nazi rule. *See, e.g.*, HERBERT, *supra* note 222, at 121 (7.8 million forced laborers and 500,000 slave laborers/concentration camp inmates in August 1944); SPOERER, *supra* note 80, at 9 (7.6 million forced laborers and 400,000 camp inmates); PAWLITA, *supra* note 43, at 15 (7.8 million forced laborers in August 1944).

[308] *See, e.g.*, MICHAEL EHRMANN, DIE GESCHICHTE DES WERKES SINDELFINGEN DER DAIMLER-MOTOREN–GESELLSCHAFT UND DER DAIMLER-BENZ AG (1998); NEIL GREGOR, DAIMLER-BENZ IN THE THIRD REICH (1998); BERND GREINER, "IG-JOE": IG FARBEN-PROZESS UND MORGENTHAU-PLAN (1996); MOMMSEN & GRIEGER, *supra* note 27; HANS POHL ET AL., DIE DAIMLER-BENZ AG IN DEN JAHREN 1933 BIS 1945: EINE DOKUMENTATION (1986).

[309] *See* HANS MOMMSEN, THE LEGACY OF THE HOLOCAUST AND GERMAN NATIONAL IDENTITY 5–6 (The Leo Baeck Memorial Lecture 42, 1999) ("More and more the Holocaust becomes the central reference point for any critical evaluation of the Nazi system at large. In addition to that, the generational change accelerated the sweeping process of lifting hitherto well established taboos regarding those aspects of Nazi history which formerly had been put into oblivion because of their actual and painful implications.").

This is not to say that historical evidence has been wholly unavailable until recently. As historian Ulrich Herbert recently noted, "[t]he claims were known since 1945."[310] While some believed that the public learned about the labor program as a result of the suits in the United States, research reveals a complex web of biography, historiography and politics.[311] We cannot exhaustively review the historical evidence establishing the relative responsibility of the state and private sector for the Nazi labor program in this Article. Instead, our goal is to review the common assertions regarding the "fateful entanglement" (*"schicksalshafte Verstrickung"*)[312] of German industry in the employment of slave and forced laborers. The defensive posture that corporate actors were involuntarily mixed up in a web of essentially war-related, state-run activity is reflected in the Foundation Law, notwithstanding that this conception has always rested on shaky evidentiary grounds. Nevertheless, even poorly founded refutations of plaintiffs' allegations won the day initially because of the political climate that existed when the earliest claims were made. Claims heard in post–World War II Germany during a period of economic reconstruction prepared the ground for an altogether superficial set of judicial decisions.[313]

In 1951, when Norbert Wollheim sued I. G. Farben for DM 10,000 for injuries arising out of his enslavement at an I. G. Farben plant at Auschwitz,[314] the defendant company argued that "Wollheim had been neither beaten nor injured, and that whatever happened to him was the responsibility of the SS, the Nazi party, the State, the subcontractors, or possibly the corrupt inmates themselves."[315] The company further claimed that "it tried to improve conditions of the workers by providing a supplementary soup . . . ."[316] Indeed, the company asserted, "if the in-

---

[310] HERBERT, *supra* note 272, at 558 (trans. by Zumbansen).

[311] *See* MOMMSEN, *supra* note 309, at 5 (attributing the newly gained international recognition of Third Reich research by German historians mainly to paradigmatic changes in the perception, centering now on the Holocaust as reference point for other evaluations of the Nazi system).

[312] This defensive posture, often occurring with regard to the cooperation between the industry and state authorities in the context of forced labor, was used, for example, by Manfred Gentz, Daimler-Chrysler's Head of Finance and one of the key initiators and later the coordinator of the German Industry's Foundation Initiative, in a speech delivered in Frankfurt in December 2000. Manfred Gentz, Address in Frankfurt (Dec. 2000).

[313] None of the American lawsuits surpassed the procedural defenses and reached the so-called merits, so we draw the substantive defenses mainly from the history of these cases in German courts, merely anticipating that they would have been replayed in some form had they proceeded beyond summary judgment in United States courts.

[314] In the Wollheim case, after the plaintiff had defeated the defendant I. G. Farben in the first instance, the plaintiff received very negative press coverage. The Wollheim victory (DM 10,000) was appealed and settled before a final verdict could be issued by the Oberlandesgericht (Higher Regional Court). The settlement deliberations, aided by the Jewish Claims Conference, eventually brought about a compensation sum of DM 30 Million to be distributed to former forced laborers. *See* Benz, *supra* note 258.

[315] FERENCZ, *supra* note 8, at 36.

[316] *Id.* The soup referred to is the notorious "Buna soup," named for the synthetic rub-

mates had not been employed they would have been killed even sooner. The implication was that Wollheim should have been grateful to [I. G.] Farben that he was still alive."[317]

Corporate leaders usually generate three interrelated substantive defenses when confronted with claims by former laborers. The first is that the use of slaves was justified on grounds of *necessity*;[318] in using slave and forced laborers in their production processes, the companies argued, they were acting at the state's behest and without the freedom to refuse.[319]

The second substantive defense is that in comparison to the conditions the workers were facing in the death camps, the workplace was a better and safer place to be.[320] This argument was generally offered immediately after a plaintiff had established the active participation of corporate actors in the forced labor program.[321] It was deployed mainly to thwart claims based on abusive treatment. While research shows that the "supplementary soup" was devoid of nutritional value and that foremen often amused themselves by kicking over the soup pot, the supplementary soup argument successfully defeated allegations of maltreatment in German courts.

The third defense can be called the *Schindler defense*.[322] This line of argument maintains that by accepting the laborers as their charges, the companies actually saved their lives. As to slave laborers in particular, the Schindler defense ought to be tough to make out, considering only about five percent of slave laborers survived the war.[323] It has, nonetheless, been an industry mainstay.

What is critical about these defenses for purposes of our argument is not their factual plausibility. To rebut them conclusively would require a

---

ber produced at Farben's Auschwitz facility. *See* BORKIN, *supra* note 39, at 125.

[317] FERENCZ, *supra* note 8, at 36.

[318] *See* Bazyler, *supra* note 9, at 194, 208; Brown, *supra* note 15, at 583.

[319] *See* BGH, RzW, 26 (1963), 525; BGH, NJW, 19 (1973), 1549; in this vein Ernst Féaux de la Croix, *Schadensersatzansprüche ausländischer Zwangsarbeiter im Lichte des Londoner Schuldenabkommens*, 13 NJW 2268 (1960), 2269. *See also* FERENCZ, *supra* note 8, at 109 (regarding the use of this defense by German electric companies and the acceptance of this defense by the Berlin appellate court).

[320] This line ignores the hierarchy among the laborers; while the work conditions were better for some, they were just as terrible for others. *See* SPOERER, *supra* note 80, at 90; HERBERT, FREMDARBEITER, *supra* note 52, at 17; CHRISTOPHER R. BROWNING, NAZI POLICY, JEWISH WORKERS, GERMAN KILLERS 93 (2000) (concerning Jewish workers in Poland specifically).

[321] SPOERER, *supra* note 80, at 96, and others have shown that corporate actors not only applied to the respective authorities for forced laborers, but that they actively formed joint ventures with the authorities to organize the deportation, the incarceration and the exploitation of workers they selected at the ramps, *see* FERENCZ, *supra* note 8, at 1–32.

[322] *See* THOMAS KENEALLY, SCHINDLER'S LIST (1993) and Stephen Spielberg's movie SCHINDLER'S LIST (Universal Pictures 1994), telling the story of a heroic German industrialist who saved the lives of Jews who would have been sent to the gas chamber but for his willingness to pay the Nazis for their labor. See also FERENCZ, *supra* note 8, at 190.

[323] *See* Imre Karacs, *Germany's Pounds 3.2BN Bid to Close Book on Nazi Past Leaves only Rifts and Rancour in its Wake*, INDEPENDENT, July 18, 2000, at 3.

mountain of historical evidence that would vary from company to company.

Instead, what is important is what a finder of fact or law would have to consider in order to determine the extent to which the evidence established the companies' legal responsibility for the use of slaves. It is not enough for a judge or jury to determine that the lost and starving spelunker in Lon Fuller's famous parable killed and ate his companion; to determine whether the defendant is guilty of murder, someone must determine whether the circumstances justified it.[324] Any factual finding would have to be followed by what amounts finally to a moral judgment.

The evidence may show a requisition form ordering a new batch of laborers;[325] or the payment of fees to the regime for the slaves; or abuse of slaves by a company's German employees; or inhumane living conditions. Whether these facts establish the companies' liability, however, requires something more. The necessity defense in particular requires a finding that the companies had no choice, but to decide whether the companies had a choice, the decision-maker will have to decide what constitutes a choice. Does, for example, a company's failure to defy production quotas establish its responsibility? This is a moral question as well as a legal one.

Moreover, that a moral question underlies or is coextensive with a legal question is no less true of the procedural defenses than it is of these substantive defenses. As the court proceeds through issues of timeliness and justiciability, it will deliberate on some dimension of the ultimate question. By determining whether the claims arose out of war and whether the companies acted as agents of the Reich, courts must implicitly determine whether the companies bear responsibility for their participation in the Nazi labor program. Courts are forced to confront the immense questions raised by the Holocaust, including what causes such a calamity and who fairly can be held responsible for it.

The strategy of separating moral or political responsibility from legal responsibility has served the companies well. Today's "enlightened German industrialists"[326] have earned endless congratulations for "courageously com[ing] to terms with injuries largely ignored for 55 years,"[327] while at the same time leaving themselves room to deny the validity of the legal claims. A close examination of the legal issues reveals, however, that the separation is a false one, and that some of the most difficult moral questions associated with the Holocaust—questions about agency and about why it happened—underlie the particular legal issues raised. It

---

[324] *See* Lon L. Fuller, *The Case of the Speluncean Explorers*, 62 HARV. L. REV. 616 (1949).

[325] *See* FERENCZ, supra note 8, at 189.

[326] Marilyn Henry, *Germany Finalizes Slave Labor Fund*, JERUSALEM POST, July 18, 2000, at 1.

[327] *Id.* at 12.

is impossible to make a legal claim in this context without making a moral one as well.

Ultimately, therefore, the refrain offered by the companies that they owe a moral responsibility but not a legal one is untenable. Each of the defenses that the companies proffered implies not just a disclaimer of legal liability, but also a larger denial of remorse.

Stuart Eizenstat may have been right that the claimants' chances of success in the courtroom were poor, but this cannot be said to be the unfortunate byproduct of dry, technical rules. It is the result of a mountain of political and moral judgments that have coalesced over the course of more than a half-century, even as the historical context has changed – judgments that relieve German industry of far more than just legal responsibility.

Predictions of the plaintiffs' failure did much of the work of directing the claims toward their diplomatic and legislative resolution. In the next Part, however, we argue that a potent rhetorical strategy buttressed the impression that litigation was ill-suited to addressing the slave and forced labor claims and deflected criticism of the compensation amounts.

IV. THE RHETORIC OF REVERENCE: IS THE HOLOCAUST INCOMPATIBLE WITH LITIGATION?

Supporters of the Foundation Law have justified diversion of legal claims from judicial to diplomatic and legislative fora by reference to the sanctity of the Holocaust. The claims of Holocaust survivors, the argument goes, deserve a more dignified setting than the courtroom, where lawyers quibble over evidentiary and procedural rules for the sake of pecuniary gain. Such assertions, though couched in reverence for the Holocaust, flip reverence on its head, disadvantaging victims in their quest to obtain legal redress.

Critics of the Holocaust-related lawsuits have charged that "[t]he pursuit of billions in Holocaust guilt money has gone from the unseemly to the disgraceful . . . a treasure hunt for hungry tort lawyers and major Jewish organizations."[328] To such critics, "[t]he Holocaust commands the preservation of memory . . . [b]ut contingency fees? Class-action suits? . . . Is this what honoring the Holocaust has come to? A shakedown"?[329] Commentator Charles Krauthammer went so far as to say "[T]o focus memory on money is literally to debase the sacred."[330]

Norman Finkelstein gained international attention by taking a particularly flashy whack at the claims: "[t]he current campaign of the Holocaust industry to extort money from Europe in the name of 'needy Holo-

---

[328] Charles Krauthammer, *The Holocaust Scandal*, WASH. POST, Dec. 4, 1998, at A29.
[329] *Id.*
[330] *Id.*

caust victims' has shrunk the moral stature of their martyrdom to that of a Monte Carlo casino."[331] Finkelstein might be regarded as a fringe commentator, but long-time national director of the Anti-Defamation League Abe Foxman had similar things to say about the litigation efforts, derisively calling the lawsuits "[a] new 'industry'"[332] and the "opportunity of a lifetime"[333] for lawyers.

Further, some commentators could not help but worry about the "revival of Shylockian stereotypes"[334] that the claims were sure to evoke. Incredibly, even Foxman, at the helm of what is perhaps the leading organization in the fight against anti-Semitism, expressed anxiety that "all the talk about Holocaust-era assets is . . . making the century's last word on the Holocaust that the Jews died not because they were Jews, but because they had bank accounts."[335] Then, (as if to establish the continuing saliency of anti-Semitic imagery), Foxman disparaged one of the more high-profile lawyers associated with the lawsuits, Ed Fagan, calling him an "ambulance chaser," who has no place "in this serious and sacred undertaking."[336]

---

[331] NORMAN G. FINKELSTEIN, THE HOLOCAUST INDUSTRY 8 (2000).

[332] Abraham Foxman, *The Dangers of Holocaust Restitution*, WALL ST. J., Dec. 4, 1998, at A18.

[333] *Id.*

[334] Krauthammer, *supra* note 328, at A29 (referring to the stereotypical Jewish character Shylock in William Shakespeare's *A Merchant of Venice*).

[335] Foxman, *supra* note 332, at A18. *But see* Richard Cohen, *The Money Matters*, WASH. POST, Dec. 8, 1998, at A21 (arguing that "there's something perverse for Jews to have their money or property taken and then to have to worry about being called cheap or avaricious for demanding it back," and that while evoking anti-Semitic images is a risk, "it is worse to implicitly honor the stereotype by refraining from doing what others would do as a matter of course—including suing for damages.") *Id.*

[336] Foxman, *supra* note 332, at A18. This sentiment was echoed by Melvyn Weiss, another of the claimants' attorneys who called Fagan "a relatively small-time lawyer who thought himself bigger than he is." Steve Chambers, *Holocaust Lawyer Inspires Range of Emotions*, STAR-LEDGER, Dec. 20, 1998, at 27.

Perhaps Foxman was unaware of the dubious history of the ambulance chaser archetype in American law. As Deborah Rhode has observed, "[a]t the close of the nineteenth century, the recently-founded American Bar Association . . . began spearheading a campaign for higher professional standards. While the quest was 'aimed in principle against incompetence, crass commercialism, and unethical behavior,' the ostensibly 'ill-prepared' and 'morally weak' candidates were often . . . 'of foreign parentage, and, most pointedly, Jews.'" Deborah L. Rhode, *Moral Character as a Professional Credential*, 94 YALE L.J. 491, 500 (1985) (citing MAGALI SARFATTI LARSON, THE RISE OF PROFESSIONALISM 173 (1977)). The ethical canons governing solicitation were drafted as a rebuke and prohibition "against those in personal injury practice, who bore the pejorative label 'ambulance chasers.'" JEROLD S. AUERBACH, UNEQUAL JUSTICE: LAWYERS AND SOCIAL CHANGE IN MODERN AMERICA (1976) *in* DEBORAH L. RHODE & DAVID LUBAN, LEGAL ETHICS 695 (1992). Auerbach writes that "[n]ot only were [the so-called ambulance chasers] criticized for professional malfeasance; their speech was mocked (many were recent immigrants) and their perseverance was denigrated as aggressiveness (many were Jewish)." *Id.* at 696. "The ethical crusade that produced the Canons concealed class and ethnic hostility [toward] Jewish and Catholic new-immigrant lawyers of lower-class origin." *Id.* This hostility was expressed, in part, in the term "ambulance chaser."

Reverberations of this notion that the memory of the Holocaust is debased by the pursuit of money damages can be heard in some of the defensive posturing by Foundation proponents who can frequently be heard to say that the claims were "not about the money" but rather about memory or acknowledgement. Gideon Taylor of the Jewish Claims Conference described his own job as

> fundamentally impossible . . . . You're taking the greatest moral challenge, the Holocaust, and putting it together with the most base form of human interaction—money. You can't really do compensation, you can't be made whole. So I speak of a measure of justice, something that's symbolic, that helps survivors live with a measure of dignity.[337]

The message from German industry sounds strikingly similar. This from Volkswagen's marketing director, Klaus Kocks: "for us . . . this is not about money. But what is adequate? 400 German Marks? 600 German Marks? 4.000 German Marks? The questions contain an inner cynicism that is outright irritating."[338]

The Law's preamble confirms this sentiment. "[T]he injustice committed and the human suffering it caused cannot be compensated by financial payments."[339]

The alleged incompatibility of the memory of the Holocaust with money also gave rise to a controversy regarding the propriety of attorneys' fees. In March 1999, the American Gathering of Jewish Holocaust Survivors passed a resolution admonishing attorneys to perform their work *pro bono* on the grounds that charging fees "'demeans the rights and memories' of Holocaust victims and survivors."[340] The executive director of the World Jewish Congress agreed, saying, "Our view is that lawyers should take this on pro bono."[341] After the conclusion of negotiations resulting in the Foundation, the *New York Times* opened its report on lawyers' fees this way: "Lawyers who represent Nazi-era slave laborers split more than $52 million in legal fees yesterday for work on a case that will bring Holocaust victims $5,000 to $7,500 each."[342]

---

[337] Jan Hoffman, *Public Lives: Nazi Reparations Require Attention to Detail*, N.Y. TIMES, June 16, 2000, at B2.

[338] Evers, *supra* note 14, at 225–26. Eizenstat somehow managed to answer Kocks's question without committing the offense of which Kocks speaks with the strange phrase "dignified sum." *Eizenstat Testimony on Worldwide Holocaust Restitution Efforts: Hearing before the Senate Foreign Relations Comm.*, 106th Cong. (2000) (statement of Eizenstat, U.S. Deputy Sec'y and Special Rep. for Holocaust Issues).

[339] Law on the Creation of the Foundation, Preamble, v. 2.8.2000 (BGBl. I S.1263).

[340] Elissa Gootman, *Holocaust Survivors Criticize Lawyers for High Fees*, FORWARD, Mar. 26, 1999, at 4.

[341] *Id.*

[342] Jane Fritsch, *$52 Million for Lawyers' Fees in Nazi-Era Slave Labor Suits*, N.Y. TIMES, June 15, 2001, at A10. The article reported that because "it would be unseemly to

We can understand why one might experience unease with the prospect of compensating survivors of such an immense tragedy as the Holocaust with dollars. Surely Taylor was right when he denied the possibility that money could ever make whole the victims of as bestial a program as that of the Nazis.' "[W]e . . . experience the universality of money as somehow degrading," observes Michael Walzer.[343] Sometimes "money fails to represent value; the translations are made, but as with good poetry, something is lost in the process."[344]

Money damages is the way of a legal system that can only do so much. We may lament, but we do not protest, the use of money damages to compensate, for example, the parent of a child struck and killed by a drunk driver. Still, "if money answereth all things, it does so . . . behind the backs of many of the things and in spite of their social meanings."[345]

Anxiety about trivializing the Holocaust by monetizing it is not a new feature of Holocaust debate. The first post-war German chancellor, Konrad Adenauer, secretly initiated compensation talks with representatives of the Israeli government and Jewish organizations in the face of widespread resentment from the German public.[346] Israel was torn apart by the emotionally charged dilemma over whether to accept the financial restitution he offered.[347]

Moreover, it is not only money's association with the Holocaust that has provoked impassioned discord. It seems that any analogy to, or any categorization of, the Holocaust can be experienced as degrading. Historians and social commentators spar obstinately over whether the Holo-

---

pay the lawyers the traditional contingency fee of one-third[,] . . . [t]he lawyers agreed to split a pool of money from 1 to 1.5 percent of the total." *Id.* (quoting E. Stuart Eizenstat, the American Government's chief negotiator).

[343] MICHAEL WALZER, SPHERES OF JUSTICE: A DEFENSE OF PLURALISM AND EQUALITY 97 (1983).

[344] *Id.*

[345] *Id.* at 103.

[346] WIEDERGUTMACHUNG UND KRIEGSFOLGENLIQUIDATION, *supra* note 43, at 106–07. For critical accounts of the German political context of these compensation talks, *see, e.g.*, FREI, *supra* note 272; REICHEL, *supra* note 270, at 73–74; Yeshayahu A. Jelinek, *Israel und die Anfänge der Shilumim*, *in* WIEDERGUTMACHUNG IN DER BUNDESREPUBLIK DEUTSCHLAND 123 (Ludolf Herbst & Constantin Goschler eds., 1989); Benz, *supra* note 258, at 138–40; Constantin Goschler, *Streit um Almosen: Die Entschädigung der KZ-Zwangsarbeiter durch die deutsche Nachkriegsindustrie*, 2 DACHAUER HEFTE 175 (1986).

[347] *See* WIEDERGUTMACHUNG UND KRIEGSFOLGENLIQUIDATION, *supra* note 43, at 107, 123–24.

caust is "unique"[348] and "unrepresentable."[349] Lucy Dawidowicz, a leading historian on the subject, insists "[t]he fate of the Jews under National Socialism was unique."[350] She continues, "[A]ll too often the necessary and essential distinction between the murder of the six million Jews and the accelerating violence and terror of our time is blurred, sometimes erased, whether mindlessly or with political intent."[351] Elie Wiesel also has maintained the uniqueness of the Holocaust, as well as its essential mystery, warning that no one who did not experience it could ever begin to understand it or "transform [it] into knowledge."[352]

---

[348] *See, e.g.*, INGA CLENDINNEN, READING THE HOLOCAUST 10–12 (1999); LUCY S. DAWIDOWICZ, THE HOLOCAUST AND THE HISTORIANS 15–19 (1981); PETER NOVICK, THE HOLOCAUST IN AMERICAN LIFE 9, 14–15 (1999). While some who have engaged in this debate have done so by going through the rather lawyerly exercises of analogizing the Holocaust to and distinguishing it from other historical tragedies, Novick takes a more critical approach, looking to the political meaning of uniqueness arguments. In its typical usage, Novick dislikes the uniqueness position, calling it "deeply offensive. What else can [it] mean [he asks] except 'your catastrophe, unlike ours, is ordinary; unlike ours is comprehensible; unlike ours is representable.'" *Id.* at 9. Novick remains focused, however, on the political context in which uniqueness is invoked. For instance, he recalls

> the insistence of Chancellor Helmut Kohl's party that as a price for supporting the law against denying the Holocaust, the law had to include a provision making it illegal to deny the suffering of Germans expelled from the East after 1945. In this German context—and context, as always, is decisive—"relativization" meant equating crimes *against* Germans to crimes *by* Germans . . . . Those Germans who insisted on the uniqueness of the Holocaust, who condemned its relativization, did so to block what they correctly regarded as a move to evade confrontation with a painful national past . . . . The identical talk of uniqueness and incomparability surrounding the Holocaust in the United States performs the opposite function: it promotes *evasion* of moral and historical responsibility [for wrongs committed against] blacks, Native Americans, Vietnamese, or others . . . .

Id. at 14–15.

[349] See, for example, PROBING THE LIMITS OF REPRESENTATION: NAZISM AND THE "FINAL SOLUTION," (Saul Friedlander ed., 1992) for a range of perspectives on whether the Holocaust defies all of our pre-existing modes of expression and analysis. As Clendinnen explains, "we expect the magic of art to intensify, transfigure and elevate actuality. Touch the Holocaust and the flow is reversed. The matter is so potent in itself that when art seeks to command it, it is art which is rendered vacuous and drained of authority." CLENDINNEN *supra* note 348, at 164–65. Fiction readers, for example, "typically expect to be seduced into concern for particular characters, who are then pursued through time and different contexts to some plausible and emotionally satisfying outcome . . . . [But when the story regards the Holocaust], *any* good outcome, *any* act of dignity or defiance, appears as a falsification or sentimentalisation of the general condition." *Id.* at 167–68.

[350] DAWIDOWICZ, *supra* note 348, at 11.

[351] *Id.* at 16. *See also* LUCY S. DAWIDOWICZ, THE WAR AGAINST THE JEWS 1933–1945 xxxvi (1975) ("The Final Solution transcended the bounds of modern historical experience.").

[352] CLENDINNEN, *supra* note 348, at 11, 21 (citations omitted). Clendinnen finds Wiesel's view "abstract and remote," more "theological" than concerned with "the human." *Id.* at 44.

Stuart Eizenstat, on the other hand, seems to find Wiesel's view authoritative: "No one can exceed his . . . towering moral stature on this subject." *Eizenstat Testimony on Worldwide Holocaust Restitution Efforts: Hearing before the Senate Foreign Relations Comm.*, 106th Cong. (2000) (statement of Eizenstat, U.S. Deputy Sec'y and Special Rep. for Holo-

The refusal to analogize to the Holocaust likely derives from the understandable desire to secure the appropriate degree of awe and respect, but defying analogy is precarious when it comes to law. In his captivating article, *Film as Witness*,[353] Lawrence Douglas explains how the prosecutors at Nuremberg "anticipated the crisis of representation that has come to characterize efforts to . . . captur[e] the Holocaust's central horror."[354] The need to "establish incredible events by credible evidence"[355] prompted prosecutors to turn to film rather than rely on eyewitness testimony to describe the camps. Douglas's article studies the

> consequence[s] of an attempt to comprehend an unprecedented evil through an idiom whose authority is anchored in the concept of precedent—and in the belief that all crimes can and must be judged according to familiar principles filtered through past experience.[356]

Douglas demonstrates that the very horror that so demanded prosecution of the Nuremberg defendants acted as an obstacle to proving the defendants' guilt.

The film *Music Box*[357] dramatizes a quite similar point.

> [t]he film's action centers on the trial in the United States in the late 1980's of a man accused of having perpetrated atrocities in Hungary in 1944–45. A quietly dressed woman, reserved in a manner as middle-class European women tend to be, is in the witness box. She is saying incredible things, things she seems to find incredible herself even as she is saying them, about the vile and most carnal connection there had been, one dreadful day, between herself and the impassive old man seated opposite her.
>
> Her listeners are outraged by what they are hearing; they are angry, or they weep. But they are also bewildered. They know that what they are hearing is true. It remains, nonetheless, incredible. Imagination makes its accustomed leap—and falls. The difficulty does not lie in the representation: the woman's

---

caust Issues), at 1.

[353] Lawrence Douglas, *Film as Witness: Screening Nazi Concentration Camps Before the Nuremberg Tribunal*, 105 YALE L.J. 449 (1995).

[354] *Id.* at 452.

[355] *Id.* (citing a report by the United States Supreme Court Justice Robert H. Jackson in his capacity as chief prosecutor).

[356] *Id.* at 453. Douglas argues that this dilemma explains the prosecution's decision to use film footage that "understands the crimes to be the consequence of aggressive militarism rather than genocide." *Id.* at 477. His theory is that this conception of the crimes enabled prosecutors to harmonize shocking (unprecedented) imagery with "concepts of ordered legality" (precedent). *Id.* at 479.

[357] MUSIC BOX (Artisen Entertainment 1989).

> words are all too graphic. It is we, the listeners, who recoil, who are baffled, who are at both an imaginative and a cognitive loss .... We are persuaded of the truth of her words. It is just that we cannot believe them.[358]

The perception that the Holocaust is unique, or unprecedented, or incredible may derive of grief and respect. But in law at least, uniqueness and incredibility come with certain risks.

So it goes with sanctification, as well. Sanctification of the memory of the Holocaust does not necessarily protect Holocaust survivors in every context.

Rather than protecting or honoring the survivors, reverence for the memory of the Holocaust created a hostile atmosphere for their legal claims. The appeal to sanctity suppressed criticism of the low payment amounts available under the Foundation Law by subjecting such criticism to the charge that it is crass. The barrage of derision levied at the lawsuits and the lawyers created the impression that litigation is too vulgar a mechanism for resolving claims arising out of the Holocaust. It impelled the claims to other fora, such as diplomacy, legislation, and private distribution, all imagined to epitomize noblesse when contrasted with the dirty rough and tumble of litigation.

Finally, the defensive assertion that the claims were "not about the money," steered observers' attention to an alternative objective for the claims: *acknowledgement. It is not fitting to decry the low amounts*, the suggestion went, *because the claims were about something else altogether.*[359] The mere identification of acknowledgement as the foremost objective, however, does not establish that it has been achieved. As we argued in Part III, the companies never really acknowledged any wrongdoing.[360] They denied legal culpability from the outset, preferring instead to recognize a "moral obligation." Yet once the particular defenses they offered are scrutinized, this combination of claims seems untenable. If acceptance of the restitution amounts rests on the belief that something greater than money was gained, we should think again.

---

[358] CLENDINNEN, *supra* note 348, at 19.

[359] One could certainly object to this line of thinking on the grounds that while any one survivor might wish to adopt the position that her claim is not about the money, it is entirely a different matter for this position to be taken for her.

[360] Martin Jay's comment in an essay about literary representation of the Holocaust is applicable to the legal quest for acknowledgement: "it is precisely the distinction between those who acted and their victims that must be scrupulously retained in any responsible account of the horror of those years." Martin Jay, *Of Plots, Witnesses, and Judgments*, *in* PROBING THE LIMITS OF REPRESENTATION: NAZISM AND THE "FINAL SOLUTION," 101 (Saul Friedlander ed., 1992).

V. CONCLUSION

Had the people who sued Degussa, Ford Werke, Siemens, and other German companies proceeded through conventional litigation processes, they might have lost. They might also have died before the verdicts were delivered. This Article is not an homage to the promise of litigation to bring justice to the wronged. It is an inquiry into the termination of claims asserted in a judicial forum in favor of a public law, drafted using a diplomatic process and implementing an exclusively private distribution scheme. This move drastically limited payments and left claimants with truncated avenues to pursue relief.

This outcome might have been justified by the attainment of some other, perhaps higher, objective, but there was none. The legal practices leading to and constituting the Foundation enabled the culprits to evade accepting any real responsibility for the enslavement of millions, while cloaking this evasion in reverence for the slaves. Through careful institutional maneuvering and shrewd justificatory rhetoric, the successors to the legacy of slave and forced labor managed to create an impression of honor and noblesse, while undermining what they trumpeted as the one legitimate objective of the process.

ARTICLE

ON THE ECONOMIC EFFICIENCY OF USING LAW TO INCREASE RESEARCH AND DEVELOPMENT: A CRITIQUE OF VARIOUS TAX, ANTITRUST, INTELLECTUAL PROPERTY, AND TORT LAW RULES AND POLICY PROPOSALS

RICHARD S. MARKOVITS*

This Article delineates a third-best distortion-analysis approach to analyzing the allocative efficiency of any private or public choice and executes partial and preliminary third-best-allocative-efficiency ("TBLE") analyses of a variety of actual and proposed policies related to research and development ("R&D"). TBLE analyses are distinguished by two facts: (1) they take into account the reality that the impact of policy choices that reduce the number or magnitude of the Pareto imperfections in the system on any given type of resource misallocation will depend on the remaining imperfections in the system that would individually cause the relevant type of misallocation and (2) they take into account the inevitable cost and imperfectness of both data and theoretical analyses. The partial and preliminary TBLE analyses this Article executes call into question the current consensus that from the perspective of allocative efficiency our society devotes too few resources to investment in general and to all types of R&D in particular. The Article argues that, although we probably devote too few resources to production-process research from the perspective of allocative efficiency, we devote too many resources to product R&D (and to quality-or-variety-increasing investment in general). The Article then examines the implications of this conclusion for the allocative efficiency of various actual and proposed tax, antitrust, intellectual property, and tort policies, many of which were designed to increase investment and R&D non-selectively rather than to increase production-process R&D without increasing product R&D or quality-or-variety-increasing investment in general.

Virtually all politicians, economists, and law and economics scholars believe that we devote fewer resources to investment in general and to research and development ("R&D") in particular than is allocatively efficient.[1] This consensus has led to a related policy consensus that it would

---

* John B. Connally Chair in Law, University of Texas School of Law; B.A., Cornell University, 1963; Ph.D., London School of Economics, 1966; L.L.B., Yale University, 1968; M.A., Oxford University, 1981.

The author would like to thank Wendy Gordon, Paul Heald, Calvin Johnson, Mark Lemley, Tom McGarity, Neil Netanel, Tony Reese, and Bernd Schäfer for comments on an earlier draft of this manuscript.

[1] The economists base their conclusions on both theoretical and empirical analyses. The theoretical analyses focus primarily on externality-spillovers from knowledge-creation. The empirical analyses attempt to measure the so-called "social rate of return" to R&D

roughly speaking, 100% times the ratio of [(the difference between the economic efficiency gains generated by the use of any discovery a research-project generates) to (the allocative cost of the project)] minus 100%. For definitions of the expressions "economic efficiency" and "allocative cost," see Part I *infra*. The seminal theoretical work on the relationship between the economic efficiency and profitability of R&D is Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609 (Richard R. Nelson ed., 1962) (demonstrating that traditional causes of market failure will result in misallocation of resources to research activities, particularly to cost-reducing inventions). For some more recent theoretical work on R&D-spillovers and other Pareto imperfections that would, acting on their own, cause the profitability of R&D to diverge from its allocative efficiency, see GENE M. GROSSMAN & ELHANAN HELPMAN, INNOVATION AND GROWTH IN THE GLOBAL ECONOMY 43–111 (1991) (assessing the impact of monopoly rents from R&D gains on the public-knowledge base and the long-term development of the global economy); Charles I. Jones & John C. Williams, *Measuring the Social Return to R&D*, 113 Q. J. ECON. 1119 (1998) (establishing that social rates of return on R&D investments far exceed private returns, and concluding that market distortions have resulted in substantial underinvestment); James J. Anton & Dennis A. Yao, *Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights*, 84 AMER. ECON. REV. 190 (1994) (demonstrating that inventors can capture a sizeable share of the value of their innovation without the aid of market-distorting mechanisms); M. ISHAQ NADIRI, INNOVATIONS AND TECHNOLOGICAL SPILLOVERS 1 (Nat'l Bureau of Econ. Research, Working Paper No. 4423, 1993) (considering empirical evidence regarding the effects of R&D investment; examining the relationships among R&D investment, productivity growth, and the rate of return on R&D investment; and discussing the magnitude of benefits from others' R&D efforts, and how these benefits are transferred); Zvi Griliches, *The Search for R&D Spillovers*, 94 SCANDINAVIAN J. ECON. 29 (Supp. 1992) (reviewing the basic model of R&D spillovers and the empirical evidence concerning their existence and magnitude, concluding that the spillovers have important effects on economic output and growth); Jeffrey I. Bernstein & M. Ishaq Nadiri, *Research and Development and Intra-Industry Spillovers: An Empirical Application of Dynamic Duality*, 56 REV. ECON. STUD. 249 (1989) (maintaining that intra-industry R&D spillovers affect firms within the industry by reducing costs, causing factor demand changes, and affecting capital accumulation rates); Philippe Aghion & Peter Howitt, *A Model of Growth Through Creative Destruction*, 60 ECONOMETRICA 323 (1992) (arguing that intertemporal spillover effects of research that causes obsolescence create distortion effects that can lead to non-optimal growth levels); Nancy T. Gallini & Brian D. Wright, *Technology Transfer Under Asymmetric Information*, 21 RAND J. ECON. 147 (1990) (developing a model for the relationship between technology innovators and licensees, which is used to demonstrate (1) how information asymmetry and the ability of the licensee to "invent around" technology can influence the terms of license agreements, and (2) how some economic rents may thus go to the licensee); Giovanni Dosi, *Sources, Procedures, and Microeconomic Effects of Innovation*, 25 J. ECON. LIT. 1120 (1988) (analyzing the processes that lead firms from potential technological opportunities to realized innovative efforts and, finally, to changes in the structures and performance of industries); Adam B. Jaffe, *Technological Opportunity and Spillovers of R&D: Evidence From Firms' Patents, Profits, and Market Value*, 76 AM. ECON. REV. 984 (1986) (analyzing the effects of technological opportunity and spillover R&D from other firms on the productivity of firms' R&D, and arguing that firms adjust the technological composition of their R&D in response to technological opportunity); Nancy T. Gallini & Ralph A. Winter, *Licensing in the Theory of Innovation*, 16 RAND J. ECON. 237 (1985) (arguing that in a duopoly the availability of licensing encourages research when firms' initial production technologies are similar in costs and discourages research when the firms face different initial costs); Nancy T. Gallini, *Deterrence by Market Sharing: A Strategic Incentive for Licensing*, 74 AM. ECON. REV. 931 (1984) (arguing that an established firm may license its production technology in order to reduce the incentive facing a potential entrant to develop its own, possibly better, technology); Partha Dasgupta & Joseph Stiglitz, *Uncertainty, Industrial Structure, and the Speed of R&D*, 11 BELL J. ECON. 1 (1980) (arguing that market competition leads to more investment in R&D and increases innovative devel-

opments); M. Therese Flaherty, *Industry Structure and Cost-Reducing Investment*, 48 ECONOMETRICA 1187 (1980) (analyzing how cost-reducing investments create unequal market shares within an industry); Ward Bowman, *The Incentive to Invent in Competitive as Contrasted to Monopolistic Industries*, 20 J. L. & ECON. 227 (1977) (arguing that the correlation between industry competitiveness and the incentive to innovate has been exaggerated); Morton I. Kamien & Nancy L. Schwartz, *Market Structure and Innovation: A Survey*, 13 J. ECON. LITERATURE 1 (1975) (surveying studies on the relationship between firm size and R&D investment return); Sheng Chen Hu, *On the Incentive to Invent: A Clarificatory Note*, 16 J. L. & ECON. 169 (1973) (arguing that industries that derive the most benefit from cost-reducing inventions will not necessarily have the most incentive to invest, because an industry with higher total returns may face lower marginal returns, and may therefore allocate fewer resources to R&D); Jack Hirshleifer, *The Private and Social Value of Information and the Reward to Inventive Activity*, 61 AM. ECON. REV. 561 (1971) (arguing that in the context of technological uncertainty, the acquisition and dissemination of technological information may be motivated by the ability to capture the pecuniary effects of wealth redistribution due to price revaluations that result from the release of the new information); Harold Demsetz, *Information and Efficiency: Another Viewpoint*, 12 J. L. & ECON. 1 (1969) (arguing that the problem of how to efficiently allocate resources to the production of information is best addressed by an approach in which the relevant choice is between alternative real, institutional arrangements, rather than between an ideal norm and an existing, imperfect, institutional arrangement).

The empirical literature includes F. M. SCHERER, INNOVATION AND GROWTH: SCHUMPETERIAN PERSPECTIVES (1984) (using Joseph Schumpeter's work to argue that technological change will have more of an impact on material well-being than static resource allocation); Jones & Williams, *supra*; Zvi Griliches, *Productivity, R&D, and the Data Constraint*, 84 AM. ECON. REV. 1 (1994) (arguing that a combination of nonmeasurement and mismeasurement of the effects of R&D on productivity levels prevents economists from fully understanding productivity growth and its recent slowdown); Zvi Griliches, *The Search for R&D Spillovers*, 94 SCANDANAVIAN J. ECON. 29 (Supp. 1992) (arguing that R&D spillovers are both prevalent and important, despite the econometric difficulties in measuring these spillovers); Zvi Griliches & Frank Lichtenberg, *Interindustry Technology Flows and Productivity Growth: A Reexamination*, 66 REV. ECON. & STAT. 324 (1984) (arguing that despite the substantial interindustry relationship between total factor-productivity growth and R&D intensity, the evidence for R&D spillovers remains tenuous); Zvi Griliches & Frank Lichtenberg, *R&D and Productivity Growth at the Industry Level: Is There Still a Relationship?*, *in* R&D, PATENTS AND PRODUCTIVITY 465 (Zvi Griliches ed., 1984) (demonstrating a strong relationship between the intensity of private R&D expenditures and subsequent total factor-productivity growth); F. M. Scherer, *Inter-Industry Technology Flows and Productivity Growth*, 64 REV. ECON. & STAT. 627 (1982) (concluding that empirical data fail to show that the 1970s productivity slump stemmed from a decrease in the marginal productivity of R&D); Leo Sveikauskas, *Technological Inputs and Multifactor Productivity Growth*, 63 REV. ECON. & STAT. 275 (1981) (exploring the effect of technological inputs on productivity growth, and finding support for the idea that R&D leads to higher returns than does an increase in technical knowledge); Nestor E. Terleckyj, *Direct and Indirect Effects of Industrial Research and Development on the Productivity Growth of Industries*, *in* NEW DEVELOPMENTS IN PRODUCTIVITY MEASUREMENT AND ANALYSIS 359 (John W. Kendrick & Beatrice N. Vaccara eds., 1980) (testing the extent to which the effect of privately financed R&D, which had previously been considered an increase in productivity growth, is independent of other inputs, including human capital).

As the rest of this Article will reveal, the assumption that we currently devote too few resources to R&D of all types from the perspective of economic efficiency is evident in the policy work that economists and law and economics scholars do. Thus, as Merges and Nelson have indicated, every effort to design an economically efficient patent system has proceeded on the assumption that, when it comes to invention and innovation, "faster is better." *See* Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 878 (1990) (analyzing the economic effects of patent scope, and arguing that patent law has and should continue to favor the race to invent). Of course, patent policy will affect economic efficiency in many ways other than by

be allocatively efficient and desirable overall to use tax, antitrust, intellectual property ("IP"), and tort law to increase the amount of resources devoted to investment in general and to R&D in particular. For example, economists, lawyers who make considerable use of economics, and an increasing number of politicians advocate relaxing pro-competition policies to permit firms to enter into mergers, joint ventures, and other types of consortia that will allegedly increase the quantity and/or proficiency of their participants' research even when such business arrangements will reduce competition.[2] In the same vein, judges and legislators have in-

---

influencing the amount of resources devoted to R&D (for example, by affecting the amount of misallocation generated by the underproduction of discovered products or the underuse of discovered production-processes, by inducing allocatively inefficient efforts to invent around awarded patents, by changing the balance between basic and applied research, by encouraging inefficiently duplicative research projects, by altering the extent of the efforts made to improve existing patents, and by affecting the amount of allocative transaction costs generated by attempts to secure, defend, evade, license, and sell existing patents). For two general analyses of the economic efficiency of the patent system (written respectively by a legal academic who makes considerable use of economics and by an economist), see Kenneth Dam, *The Economic Underpinnings of Patent Law*, 23 J. LEGAL STUD. 247 (1994); Staff of Senate Subcomm. on Patents & Copyrights, 85th Cong., *An Economic Review of the Patent System*, Study No. 15 (Comm. Print 1958) (Fritz Machlup, author) [hereinafter Machlup].

For a partial bibliography of antitrust-policy economists and lawyers whose R&D-related proposals reflect their assumption that we currently devote too few resources to R&D from the perspective of economic efficiency, see *infra* note 2.

[2] *See* Thomas A. Piraino, Jr., *Reconciling Competition and Cooperation: A New Antitrust Standard for Joint Ventures*, 35 WM. & MARY L. REV. 871 (1994) (proposing a new standard for judicial antitrust analysis of joint ventures, based on the economic benefits that can result from such arrangements); Thomas M. Jorde & David J. Teece, *Innovation and Cooperation: Implications for Competition and Antitrust*, J. ECON. PERSP., Summer 1990, at 75 (suggesting modifications to existing antitrust law aimed at promoting cooperative activity among competitors, thus motivating the development of new technologies while stimulating competition).

For arguments in favor of relaxing antitrust laws specifically in the context of high-tech industries, see Thomas M. Jorde & David J. Teece, *Innovation, Cooperation, and Antitrust: Striking the Right Balance*, 4 HIGH TECH. L. J. 1 (1989) (arguing that antitrust law impedes cooperative innovation); Janusz A. Ordover & William Baumol, *Antitrust Policy and High Technology Industries*, OXFORD REV. ECON. POL'Y, December 1988, at 13 (arguing that antitrust laws should be applied sparingly to mergers that stimulate the production and diffusion of knowledge, because coordinating R&D activities of firms is likely to provide the socially optimal amount of investment in the production of knowledge); Janusz A. Ordover & Robert D. Willig, *Antitrust for High-Technology Industries: Assessing Research Joint Ventures and Mergers*, 28 J. L. & ECON. 311 (1985) (analyzing how the standard methodology of antitrust analysis should be modified to reflect the importance of R&D and innovation as competitive market forces).

Even the few experts writing in this field who reject the case for special treatment of R&D-related efficiencies seem to accept the assumption that increases in R&D of all types would be economically efficient. *See* Joseph Brodley, *Antitrust Law and Innovation Cooperation*, 4 No. 3 J. ECON. PERSP. 97 (1990) (arguing that particularly when innovation cooperation in high technology markets includes production and marketing, serious anticompetitive risks will arise). *See also* M.I. Kamien et al., *Research Joint Ventures and R&D Cartels*, 82 AM. ECON. REV. 1293 (1992) (arguing that the sum of consumer and producer surplus is lower when companies participate in joint research ventures than when they conduct R&D on their own); K. Suzumura, *Cooperative and Noncooperative R&D in an Oligopoly With Spillovers*, 82 AM. ECON. REV. 1307 (1992) (analyzing cooperative R&D at

creased or recommended increases in the breadth and duration of the protection that IP law provides discoverers it already protects and have also extended or recommended the extension of copyright protection to additional forms of intellectual property.[3]

Unfortunately, these policy recommendations and their underlying allocative-efficiency premises are not grounded in appropriate welfare-economics analysis. The premise that we devote too few resources from the perspective of allocative efficiency to investment in general and R&D in particular is too broad-sweeping in that it fails to distinguish between (1) product R&D and investments in product quality and variety ("QV investments") on the one hand and (2) investments in production-process R&D on the other. QV investments (only some of which involve research or are innovative in any significant sense) seek to create additional and superior products, increase distributive variety or quality, or raise the average speed of delivery of some product or service throughout a fluctuating demand cycle by creating additional capacity or inventory. Investments in production-process research ("PPR") seek to reduce the cost of producing a relevant quantity of an extant product.[4]

This Article argues that the distinction between the two kinds of investment and R&D is critical for R&D and investment policy because an appropriate analysis of the way in which the relevant Pareto imperfections[5] jointly distort the profitability of product and production-process research yields the conclusion that from the perspective of economic

---

the "precompetitive stage" and concluding that when significant R&D spillovers exist, socially optimal levels of R&D are not reached).

[3] *See* Part V *infra*. As Merges pointed out in 1995, IP policy and law have seen a "shift in the conceptual baseline of legislators and judges, from a presumption of no protection to a presumption that any intangible contribution ought to be protected." Robert P. Merges, *The Economic Impact of Intellectual Property Rights: An Overview and Guide*, 19 J. CULTURAL ECON. 103, 104 (1995).

[4] I hasten to add that the concepts of "QV investments" and "PPR" are not mutually exclusive. For example, investments that are designed to create a new product that a significant number of buyers prefer to its extant alternatives and whose average total cost at some relevant output is lower than the average total cost of its extant alternatives will be simultaneously QV investments and investments in PPR.

[5] Welfare economists have demonstrated that if a certain set of so-called Pareto-optimal conditions are fulfilled, it would not be possible to make someone better off and no-one worse off by changing the way in which resources are allocated even if the relevant resource-shift could be effectuated without generating any allocative transaction costs. The eight so-called Pareto-optimal conditions are (1) perfect competition among sellers, (2) perfect competition among buyers, (3) no externalities, (4) no taxes on the margin of income, (5) actor sovereignty (each actor possesses all the information it needs to make the decision it will prefer), (6) actor maximization—no-one makes calculation errors (each actor makes the decision that the information at its disposal implies it would most prefer), (7) no problems caused by buyer surplus, and (8) the preceding conditions can be fulfilled without generating any allocative transaction costs or sacrificing any other type of economic efficiency. Pareto imperfections are departures from the Pareto-optimal conditions—imperfections in seller or buyer competition, externalities, taxes on the margin of income, actor non-sovereignty, actor non-maximization, economic-inefficiency-generating buyer surplus, and relevant transaction costs.

efficiency we currently allocate too small a proportion of our economy's resources to PPR and too large a proportion to product R&D (and QV investment in general). This conclusion implies that the investment-credit, capital-expenditure expensing, and accelerated-depreciation provisions that have been proposed and adopted, all of which provide incentives to increase R&D and investment in general without regard to the type of R&D or investment to be made, are certainly too crude from the perspective of allocative efficiency and may be misallocative on balance.

Moreover, proposals to relax the antitrust laws in relation to R&D-encouraging mergers, joint ventures, and consortia of various types are subject to the same criticism: they, too, fail to distinguish between arrangements that are designed to foster production-process R&D and product R&D. Supporters of these proposals also ignore the fact that, by reducing competition and consequently distorting factor prices, they will tend to increase the excessiveness of QV investment in general (including product R&D), increase the inadequacy of production-process research, and cause a variety of other types of misallocation unrelated to investment and R&D. Similarly, recommendations to increase the breadth and duration of patent and copyright protection and to extend such protection to additional forms of intellectual property are marred by their tendency to increase the excessiveness of the amount of resources allocated to quality-or-variety-increasing investment in general and product R&D in particular from the perspective of economic efficiency.

This Article is designed to achieve two goals and to appeal to two kinds of audiences. The first goal is to generate various conclusions about investment and R&D policy in the United States. The policymakers and some of the broader "policy audience" to whom this aspect of the Article will have the most appeal will be particularly interested in (1) its argument that antitrust and tax law should be used to decrease the proportion of our society's resources devoted to both innovative and non-innovative quality-or-variety-increasing investment and to increase the proportion of our resources devoted to both production-process research and to the production of extant products, (2) its more specific conclusions that antitrust law will tend to achieve these ends to the extent that it increases price competition and that tax law will do so to the degree that it treats the profits yielded by production-process research and unit-output production more favorably than it does the profits yielded by QV-investment creation, (3) its proposal that IP and tax law be reformed so that the profit expectations of researchers conform more closely to the *ex ante* allocative efficiency of their research efforts, (4) its explanation of the reasons for concluding that IP law will be more allocatively efficient if it gives longer and broader protection to production-process discoverers than to product discoverers, and (5) its arguments that decisions by producers not to do research designed to discover less accident-and-pollution-loss-prone production processes or product-variants whose

production and consumption are less-accident-and-pollution-loss-prone should be assessed for negligence, that the difficulty of making such assessment favors making such producers strictly liable,[6] that the negligence of many potential avoiders should be assessed by comparing the allocative rather than the private costs and benefits of the avoidance-moves they rejected, and that allocative efficiency would be increased if a system of civil fines and subsidies were added to our current compensatory tort-damages scheme.

This Article has a second, more abstract goal: it attempts to explain and exemplify how allocative-efficiency analyses should be structured and executed. This second goal should be of primary interest to law-and-economics specialists, economists in general, and legal scholars who seek to predict the allocative efficiency of particular policies or legal doctrines. Admittedly, like many other non-economics-oriented law professors, a large number of IP scholars have ignored the allocative efficiency of the proposals they are considering and have attempted to justify such conduct by arguing that, since we cannot know the number of symphonies or technologically different types of cameras that would be allocatively efficient, we cannot predict the effect of R&D-related proposals on the allocative efficiency of potential-discoverer research-decisions. In fact, however, one need not determine the allocatively efficient level of research or the allocatively efficient number of discoveries to assess the allocative efficiency of a policy that will affect R&D. In particular, if potential researchers and developers are sovereign maximizers, one will be able to predict the effect of a proposal on the amount of misallocation caused by their production-process or product R&D decisions by determining whether it will reduce the absolute difference between the certainty-equivalent[7] profitability and *ex ante* allocative efficiency of marginal research-decisions of the relevant type. If potential researchers are sovereign maximizers, policies that equate the profitability and allocatively efficiency of such marginal research-decisions will induce them to make allocatively efficient research decisions—whatever the optimal amount of relevant R&D expenditures and the optimal quantity of relevant discoveries turn out to be.

---

[6] For a far-from-complete third-best-allocative-efficiency analysis of the allocative efficiency of shifting from negligence to strict liability, see Richard S. Markovits, *The Allocative Efficiency of Shifting from a "Negligence" System to a "Strict Liability" Regime in Our Highly Pareto-Imperfect Economy: A Partial and Preliminary Third-Best-Allocative-Efficiency Analysis*, 73 CHI.-KENT L. REV. 11 (1998) [hereinafter *Negligence Versus Strict Liability*]; Richard S. Markovits, *Monopoly and the Allocative Inefficiency of First-Best-Allocatively-Efficient Tort Law in Our Worse-Than-Second-Best World: The Whys and Some Therefores*, 46 CASE W. RES. L. REV. 313 (1996) [hereinafter *Monopoly and Tort Law*]. For a definition of "third-best-allocative-efficiency analysis," see Part I of this Article, *infra*.

[7] The "certainty-equivalent profitability" of a choice is an *ex ante* concept that equals the weighted-average-expected profitability of the choice *minus* the risk and uncertainty costs it imposes on the actor who makes it.

Unfortunately, even those economists, law and economics scholars, and legal academics who focus on the allocative efficiency and profitability of actual and proposed R&D policies usually fail to take into account various kinds of relevant general-equilibrium feedback relationships. The importance of those relationships has been emphasized by The General Theory of Second Best, which establishes that in a world in which some Pareto imperfections cannot or will not be removed one cannot assume that a policy that decreases the number and/or magnitude of the Pareto imperfections present in the system will even tend to increase economic efficiency on that account.[8] Most importantly, standard economic-efficiency analyses of R&D policies have ignored the ways in which the imperfections in price competition faced by the researcher's factor rivals, product rivals, and/or discovery customers distort the profitability of different types of R&D.[9] This Article points out this deficiency and several related deficiencies of the standard economic analysis and explains how to correct them.

This Article contains five parts. Part I defines some vocabulary and delineates a defensible, and possibly optimal, approach to analyzing the allocative efficiency of any private or public choice. Part II explains why there is good reason to believe that, taken together, the relevant Pareto imperfections result in the allocation of too many resources from the perspective of allocative efficiency to investment in quality and variety, including product R&D. Next, Parts III and IV explain why there is good reason to believe that, taken together, the relevant Pareto imperfections tend to result in the allocation of too few resources from the perspective of allocative efficiency to production-process research and to the production of units of output of existing products. Finally, Part V briefly examines the implications of the preceding allocative-efficiency conclusions for antitrust, tax, tort-and-environmental, and IP law proposals that are designed to reduce R&D misallocation.

---

[8] For the first formal statement of The General Theory of Second Best, see R.G. Lipsey and Kelvin Lancaster, *The General Theory of Second Best*, 24 REV. ECON. STUD. 11 (1956).

[9] This omission is made more puzzling by the fact that at least some economists who analyze the allocative efficiency of R&D decisions have recognized an analogous distortion in the profitability of R&D: the fact that any new discovery that renders its antecedents obsolete (a phenomenon known as "creative destruction") will tend to create allocatively inefficient profit-incentives on that account. *See, e.g.*, Aghion & Howitt, *supra* note 1, at 323–24. *See also* JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM, AND DEMOCRACY 83 (1942). For the definition of "distortion in the profitability of X," see Part I of this Article.

I should add that this deficiency undermines the empirical as well as the theoretical literature on the allocative efficiency of increasing R&D. Thus, most of the empirical studies cited in *supra* note 1 proceed on the assumption, explicitly stated by Jones and Williams, that "one does not need to make additional assumptions about . . . market structure . . . to determine the social return." Jones & Williams, *supra* note 1, at 1126–27.

I. VOCABULARY AND APPROACH TO ALLOCATIVE-EFFICIENCY ANALYSIS

Six sets of expressions or terms need to be defined. The first set consists of the expressions "increase in allocative (economic) efficiency" and "decrease in allocative (economic) efficiency." A private choice or government policy will be said to increase allocative efficiency if it gives its beneficiaries the equivalent of more dollars than it takes away from its victims. I say "equivalent" because the relevant winners and losers may not be able to capitalize the effects of the policy or choice on them. For example, an environmental policy that improves the air above a piece of land that only its owner values even after the policy is implemented will confer an equivalent-dollar gain on him (if it increases the equivalent-dollar value of his land to him) even though he cannot capitalize that gain (since the market value of his land will still be zero after the policy is adopted).[10] This definition implies that neither the moral permissibility of a choice (that is, its consistency with the relevant society's moral-rights commitments) nor its overall desirability can be predicted from its allocative efficiency. Allocatively efficient choices can violate people's rights, and rights commitments can require us to make allocatively inefficient choices. Moreover, allocatively efficient choices that do not violate anyone's rights may still be undesirable. Such choices may be undesirable because (1) from the applicable utilitarian[11] or non-utilitarian[12] perspective, the average-equivalent-dollar gained should be given a sufficiently lower weight than the average-equivalent-dollar lost to yield this result[13] or (2) from the applicable non-utilitarian perspective,

---

[10] For a discussion of how one should measure the equivalent-dollar gains and losses a policy generates and a related explanation of why the standard Kaldor-Hicks test for the impact of a choice on allocative efficiency is not only wrong but is generally biased in favor of the status quo, see Richard S. Markovits, *A Constructive Critique of the Traditional Definition and Use of the Concept of "The Effect of a Choice on Allocative (Economic) Efficiency": Why the Kaldor-Hicks Test, the Coase Theorem, and Virtually All Law-and-Economics Welfare Arguments Are Wrong*, 1993 ILL. L. REV. 485 (1993) (arguing that the monetized Kaldor-Hicks allocative efficiency test traditionally applied in analyzing the desirability of a private or government choice is flawed and generally biased in favor of the status quo and questioning the assumption that the desirability of a choice depends exclusively on its allocative efficiency).

[11] By "utilitarian," I mean a norm that implies that the moral desirability of any choice depends exclusively on its impact on the "total utility" of all creatures whose utility counts (or, in the more modern version, on the average utility of all creatures whose utility counts).

[12] The list of non-utilitarian distributional norms would include norms that value the ability of each individual to secure equal utility, equal resources, equal opportunities (of various specific types), or the sum of the allocative value of what he produced (in some contestable sense) and the dollars and equivalent-dollars he obtained through luck, gift, and bequest.

[13] When the sought-after moral-desirability conclusion can be generated in this way, it will often be more cost-effective to proceed by predicting the equivalent-dollar gains and losses the policy to be evaluated should be expected to generate, weighting those gains and losses to reflect the distributional norms to be effectuated and the facts that those norms make salient, and comparing the weighted equivalent-dollar gains with the weighted

the allocative efficiency of the choice under review is totally irrelevant to its desirability. Although the analyses that follow will focus exclusively on allocative efficiency issues, it is important to remember that policy conclusions cannot be derived solely from allocative efficiency conclusions.[14]

The second group of terms used in this Article describe the different ways in which resources can be used. This Article will concentrate on choices to increase the unit output ("UO") of one or more existing products, to create a QV investment, and to execute production-process research ("PPR").

The third set of expressions are those that refer to the various kinds of allocative inefficiency. Total-QV-investment ("ΣQV") misallocation, total production-process-research ("ΣPPR") misallocation, and total-unit-output ("ΣUO") misallocation will be said to be present in the economy to the extent that from the perspective of allocative efficiency nonoptimal amounts of resources are allocated to QV-investment creation, production-process-research execution, and unit-output production.

The fourth group of terms includes the allocative value ("LV"), allocative cost ("LC"), and allocative efficiency ("LE") of a choice. These terms will be contrasted with the private value ("PV"), private cost ("PC"), and (private) profitability ("Pπ") of the choice.

The allocative value of a marginal unit of output ("$LV_{\Delta UO}$") is the number of dollars by which we are better off when the unit in question is consumed instead of being costlessly destroyed. The allocative value of a marginal quality-or-variety-increasing ("QV") investment ("$LV_{\Delta QV}$") is the number of dollars by which we are better off when the investment is used—for example, when the new model is produced and sold as opposed to being abandoned or when the distributive outlet or capacity is operated as opposed to being costlessly destroyed. The allocative value of a marginal production-process-research project in a world in which everyone is indifferent to risk ("$LV_{\Delta PPR}$") is the weighted-average net equivalent-dollar gain it should be expected to generate by advancing the date(s) on which the discovery (or discoveries) it may yield should be expected to be made.

---

equivalent-dollar losses, than by determining directly whether the policy promotes the distributional norm to be effectuated. This conclusion reflects the fact that one will be able to execute an appropriately accurate version of the above protocol by making use of welfare-economics theorems to determine the allocative efficiency of the policy in question—to determine whether the equivalent-dollar gains it generates exceed the equivalent-dollar losses it imposes.

[14] Indeed, that is why, in this Article, the more-technical-sounding expression "allocative efficiency" is often used in place of the standard phrase "economic efficiency." For a full discussion of these issues, see Richard S. Markovits, *On the Relevance of Economic-Efficiency Conclusions*, 29 FLA. ST. L. REV. 1 (2001) (arguing that allocative-efficiency tests cannot by themselves serve as guides for optimal social policy, because they often yield results inconsistent with prescriptive moral conclusions).

The allocative cost of an economic action is its opportunity cost, that is, the allocative value that the resources the action consumes would have generated in their actual alternative uses. Take, for example, a choice to create a marginal QV investment with resources that would otherwise have been used to produce marginal units of output. In this case, the allocative cost of the marginal QV investment—$LC_{\Delta QV}$—would equal the allocative value of the units of output that were sacrificed to free the resources used to create the marginal QV investment (which would reflect not only their equivalent-dollar value to their consumers, but also the external benefits and costs[15] the consumption of the units of output in question would generate and any external benefits and costs that would be generated by the relevant units' production).

It is important to understand the relationship between the allocative cost of one actor's using a resource in a particular way and the allocative value that that resource would generate in its alternative use. Assume that there are two possible ways in which a marginal unit of some resource can be used. There is a more highly privately valued alternative resource-use ("ARU#1") and a less highly privately valued alternative resource-use ("ARU#2"). The private cost of the resource unit to the actor who wants to devote it to ARU#1 will be infinitesimally higher than the private value it would yield in ARU#2.[16] This conclusion reflects the fact that the actor who wants to devote the resource to ARU#1 will have to bid it away from the actor who would otherwise have devoted it to ARU#2. The allocative cost of using the resource for ARU#1 ($LC_{ARU\#1}$) equals the allocative value it would generate in ARU#2 ($LV_{ARU\#2}$). Thus, if the private value the resource yields the actor who allocates it to ARU#2 ($PV_{ARU\#2}$) diverges from the allocative value the resource would generate in ARU#2 ($LV_{ARU\#2}$), then the private cost of ARU#1 ($PC_{ARU\#1}$) will diverge from the allocative cost of ARU#1 ($LC_{ARU\#1}$) in the same direction and by the same amount.

The allocative efficiency of a resource-use choice equals the difference between the allocative value generated by the use and the allocative cost of the use. The profitability of a resource-use will not normally equal its allocative efficiency because (1) the private value a resource generates for its user often diverges from the allocative value his use of it generates ($PV \neq LV$), (2) the private cost of a resource to its user often diverges from the allocative cost of his using it ($PC \neq LC$), and (3) the preceding differences usually do not offset each other perfectly.

---

[15] An externality is a cost or benefit generated by an act that is not borne by or internalized to the actor in question because he does not have to pay for some resource his act consumes, such as clean air, or is not compensated for some benefit his act generates.

[16] This assumes that there are no taxes on the margin of income, no relevant government transfers (no relevant negative taxes), and no relevant monopsony rents. Such taxes, subsidies, and rents would create a divergence between the private cost of a resource to its buyer and the price its seller received for it.

The fifth set of terms to be defined are those used in distortion analysis. In distortion-analysis terminology, a private value, private cost, or profits figure is said to be "distorted" if it diverges from its allocative counterpart. More specifically, the private figure is said to be "inflated" if it exceeds its allocative counterpart and "deflated" if it is less than its allocative counterpart.

My distortion-analysis vocabulary distinguishes a variety of different types of distortions that require investigation. In addition to the aggregate distortion ("ΣD") in some private figure (the distortion generated by all relevant Pareto imperfections acting in concert), I often consider the distortion that would be caused in the private figure by a particular type of Pareto imperfection if it were the only type of Pareto imperfection present in the economy. For instance, this Article will investigate, *inter alia*, the monopoly distortion ("MD") in the profitability of a marginal QV investment ("MD($P\pi_{\Delta QV}$)"), the signed amount by which all extant imperfections in seller competition would cause the profitability of the marginal QV investment to diverge from its allocative efficiency if no other type of Pareto imperfection (such as monopsony, externalities, taxes on the margin of income, imperfections in information, failures of individuals or organizations to maximize, and problems caused by buyer surplus) were present. This Article will also examine the monopsony distortion ("MSD"), the externality distortion ("XD"), the tax distortion ("TD"), the non-sovereignty or imperfect-information distortion ("NSD"), the non-maximization distortion ("NMD"), and the buyer-surplus distortion ("BSD") in one or more private figures.

The relationship between these individual-imperfection-generated distortions and the aggregate distortion in some private figure is a complex one, which varies among resource-use types. The aggregate distortion that is relevant to a few types of resource misallocation can be expressed as the sum of the seven individual-imperfection-generated distortions just distinguished. Unfortunately, however, this is not true in most cases.

This Article focuses on the aggregate distortions in the private cost, private value, and profitability of resource-uses of different types. The aggregate distortion in the profitability of the marginal resource-use of any type ("ΣD($P\pi_{\Delta\ldots}$)") provides a great deal of information about the allocative efficiency of that type of resource-use in the prevailing circumstances. Thus, if the resource-use in question is marginal both in the sense of being last and in the sense of being infinitesimally small, the fact that its profitability is inflated implies that too many resources are allocated to that type of resource-use from the perspective of allocative efficiency (if the relevant choosers are sovereign maximizers); conversely, on the same assumptions, the fact that the profitability of a marginal resource-use of a given type is deflated implies that too few resources are allocated to this type of resource-use from the perspective of

allocative efficiency. The above conclusion follows from the identity $\Sigma D(P\pi_{\Delta}...) \equiv P\pi_{\Delta}... - LE_{\Delta}...$ and the fact that, if the relevant resource-user is a sovereign maximizer, the profitability of the marginal resource-use of any type equals zero for marginal resource-uses that are not only last but infinitesimally small.[17]

The sixth and final set of terms that needs to be defined is the set containing the expressions "first-best," "second-best," and "third-best" allocative-efficiency analysis. First-best-allocative-efficiency analyses investigate the allocative efficiency of responding to one or more specified Pareto imperfections in a particular way on the assumption that the imperfection or imperfections in question are the only Pareto imperfections in the economy. I use the acronym "FBLE" to refer to this type of analysis because it is similar to the English word "fable" and this type of analysis can fairly be said to be based on the fable that the rest of the economy is otherwise-Pareto-perfect.

Second-best-allocative-efficiency analyses investigate the allocative efficiency of responding to one or more specified Pareto imperfections in a particular way on realistic assumptions both about the other Pareto imperfections that the relevant economy contains and about the variety of ways in which resources are used in the economy in question. Unfortunately, given the fact that in the real world both data and analysis are inevitably costly and imperfect, second-best-allocative-efficiency analyses are impossible to execute and would be prohibitively expensive even if one could execute them. I use the acronym "SBLE" to refer to this type

---

[17] For example: if, in the one direction, $\Sigma D(P\pi_{\Delta}...) = +\$6$, then $LE_{\Delta}... \equiv P\pi_{\Delta}... - \Sigma D(P\pi_{\Delta}...) = -\$6$, where $P\pi_{\Delta}... = 0$. Verbally, on the above facts, the relevant marginal resource-use misallocated resources by $6 because the resource-use in question just broke even privately despite the fact that its profit-yield was inflated by $6. The total misallocation will equal the sum of (1) the misallocation the relevant Pareto imperfections generated by critically distorting the profitability of the relevant marginal decision and (2) any additional misallocation of the relevant type they generated by critically distorting the profitability of one or more intra-marginal decisions of the kind in question, i.e., by making it profitable to produce units other than the marginal unit whose production was also allocatively inefficient.

If, in the other direction, $\Sigma D(P\pi_{\Delta}...) = -\$6$, then $LE_{\Delta}... = +\$6$ since the relevant marginal resource-use broke even privately despite the fact that its profitability was deflated by $6. Although the marginal resource-use was not misallocative, the imperfections that caused the relevant aggregate distortion probably misallocated resources by deterring additional resource-uses of the relevant type, i.e., resource-uses whose profitability was negative despite the fact that their allocative efficiency was positive.

Admittedly, the fact that $\Sigma D(P\pi_{\Delta}...) \neq 0$ does not guarantee the presence of misallocation when the last resource use is lumpy or incremental. Thus, since $P\pi_{\Delta}...$ may be positive (e.g., may equal [+$11]) when the last decision of the relevant type is lumpy, the fact that $\Sigma D(P\pi_{\Delta}...)$ is positive (e.g., equals [+$6]) is consistent with the last decision's increasing economic efficiency (e.g., by $5 on the facts specified in the preceding parentheticals). Nevertheless, even when the last resource-use of the relevant type is incremental, the likelihood of misallocation and the weighted-average extent of the type of misallocation in question will increase with the absolute value of the $\Sigma D(P\pi_{\Delta}...)$ figure for the last resource-use of the relevant type.

of analysis because it is similar to the English word "sable"—a fur that is beautiful but prohibitively expensive.

Third-best-allocative-efficiency analyses take into account only those theoretical relationships and collect only that data that it is *ex ante* allocatively efficient to consider and collect. I use the acronym "TBLE" to refer to this type of analysis because it is similar to the English word "table" and this is the type of analysis one should bring to the policy-analysis table.[18]

In my judgment, third-best-allocative-efficiency analyses of particular policies should proceed in the following way: (1) delineate every type of resource misallocation whose magnitude the policy may affect; (2) analyze the various Pareto imperfections that would individually distort the profitability of marginal resource-uses of each relevant type and generate formulae for the aggregate distortion in the profitability of each marginal resource-use of each relevant type; (3) combine existing data on and guestimates of the pre-policy magnitudes of the relevant Pareto imperfections and such other factors as the percentages of the economy's resources used in various ways to generate estimates of the pre-policy aggregate distortion in the profitability of the marginal resource-uses of each relevant type; (4) combine existing data on and guestimates of the relevant resource-users' non-sovereignty and failure to maximize to guestimate the effect of the resource-users' human errors on the amount

---

[18] Admittedly, FBLE analyses are valuable: in particular, such analyses can reveal the various types of decisions whose profitability a given imperfection will distort when the economy is otherwise-Pareto-perfect. This information can be incorporated into TBLE analyses. Complete SBLE analyses are never executed. The literature contains some partial SBLE analyses, which typically examine the implications of the presence of one or two known irremediable Pareto imperfections on the effect of reducing other imperfections on a subset of the types of misallocation the relevant policy would affect. Such partial SBLE analyses are valuable in that they reveal the way in which different types of Pareto imperfections interactively distort the profitability of particular kinds of resource-allocation decisions, information that also can be incorporated into TBLE analyses. Oliver Williamson's economic-efficiency analyses of horizontal mergers that reduce price competition and generate marginal-cost-decreasing efficiencies may be the FBLE analysis by a highly respected economist that is best known to academic lawyers who are conversant with economics. *See* Oliver Williamson, *Economies as an Antitrust Defense Revisited*, 125 U. PA. L. REV. 699 (1977); *Economies as an Antitrust Defense: The Welfare Trade-Off*, 58 AM. ECON. REV. 18 (1968). For my critique of Williamson's analysis, see Richard S. Markovits, *Second Best Theory and the Standard Analysis of Monopoly Rent Seeking: A Generalizable Critique, a "Sociological" Account, and Some Illustrative Stories*, 78 IOWA L. REV. 327, 335 n.10, 348 n.33 (1993). Steven Shavell's analysis of the connection between (1) the imperfect way in which the common law resolves cause-in-fact issues in cases in which cause in fact must be determined from evidence about the contribution the defendant's activity makes to the *ex ante* probability of a loss's occurring and (2) the economic efficiency of making proximate-cause decisions depend on whether a strict-liability or negligence standard applies in the case at hand may be the partial SBLE analysis by a highly respected economist that is best known to academic lawyers who are conversant with economics. See Steven Shavell, *An Analysis of Causation and the Scope of Liability in the Law of Torts*, 9 J. LEG. STUD. 463 (1980). For my critique of Shavell's argument, see *Negligence Versus Strict Liability*, *supra* note 6 at 118–23.

of misallocation of each type likely to be associated with any $\Sigma D(P\pi_{\Delta}...)$ figure for that type of resource-use; (5) analyze the likely impact of the various policies on the relevant distorting imperfections and human errors and generate a preliminary assessment of the allocative efficiency of the policies; (6) guestimate the possible extent to which further empirical research would increase the accuracy of the empirical estimates of the relevant Pareto imperfections and of other parameters that are relevant to the allocative-efficiency effect under investigation; (7) execute sensitivity analyses to determine the certainty-equivalent amount by which more accurate estimates of the relevant Pareto imperfections, the percentages of our economy's resources devoted to various types of uses, and the impact of the policy in question on the relevant Pareto imperfections would be likely to increase the allocative efficiency of both the policy recommendation to be made in the instant case and various other policies to whose economic efficiency this data would be relevant; (8) estimate the possible allocative cost of additional empirical research on the various relevant parameters; (9) generate conclusions about the third-best-allocative-efficiency of doing additional empirical research; (10) initiate the empirical research the preceding analysis deemed third-best-allocatively efficient (an adjective I will also use TBLE to symbolize); (11) revise one's estimates of the likely allocative benefits and costs of additional empirical research in light of the research completed and continue to collect data so long as doing so seems TBLE; and, lastly, (12) generate conclusions about the relative allocative efficiency of the policies under review by combining the TBLE data collected with (a) the theoretical analysis of the relevant aggregate distortions in the profitability of the relevant set of marginal resource-uses, (b) the relationship between the absolute magnitude of those ΣDs and the amount of resource misallocation of the relevant types present in the economy given the relevant possibilities of human error, and (c) the determinants of the effect of the relevant policies on the relevant imperfections and distorting human errors.[19]

---

[19] This Article's TBLE analyses are extremely partial and preliminary. Both the theoretical work it presents and the data on which it relies are worse than TBLE. Thus, although I will consider the relevance of a great many Pareto imperfections to the issues under consideration, I will at times ignore the theoretical relevance of some of the Pareto imperfections that will distort the private figures I am analyzing and will persistently base my conclusions on data and guestimates that would be TBLE only if the decision in question had to be made immediately.

A related point is in order. Because (roughly speaking) the set of individual Pareto-imperfection-related distortions in the profitability of the various types of decisions on which this Article will focus have different signs, I will virtually never be able to generate conclusions about the allocative-efficiency sub-optimality or supra-optimality of the incentives that motivate economic actors to make these choices without combining *a priori* work with empirical assumptions. There is very little hard evidence about the magnitude of the relevant parameters; in particular, on (1) the extent of various Pareto imperfections, (2) the percentages of the economy's resources devoted to unit-output production, QV-

This Article focuses more on demonstrating that the various consensuses described in the Introduction have not been and are not justified and on explaining the appropriate approach to analyzing the issues in question (revealing the relevance of the various Pareto imperfections that can individually distort the resource-use choices in question) than on generating policy conclusions. However, this Article will at times guestimate the relevant parameters and analyze the allocative-efficiency and policy implications of these empirical guesses.

II. The Aggregate Distortion in the Profitability of a Marginal QV-Increasing Investment ($\Sigma D(P\pi_{\Delta QV})$): A Partial and Preliminary TBLE Analysis of the Causes of Aggregate-QV-Investment (ΣQV) Misallocation

A. The Aggregate Distortion in the Private Cost of a Marginal QV Investment

The tasks that marginal QV investors perform to create their QV investments vary with the nature of the QV investment in question. When the QV investment creates a new product-variant, the relevant QV investors must conduct market research into the potential commercial profitability of various sorts of product-variants, design the new product, build and test prototypes of the new product, perhaps build a pilot plant to produce it, and conduct a basic advertising campaign necessary to prepare potential buyers for its arrival. When the QV investment creates an additional or different distributive outlet, investors must conduct market research into the commercial attractiveness of different locations and nonlocational attributes, search for available locations or outlets, build or acquire the outlet, and prepare its potential patrons for its opening. When the QV investment is in capacity or inventory, investors must construct the capacity or produce the inventory and inform potential buyers of the increase in the speed with which the investor can supply patrons at times of peak demand.

In my judgment, economic actors withdraw almost all of the resources they use to create marginal QV investments from three alternative types of uses: (1) the production of additional units of existing products, (2) the execution of production-process-research projects, and (3) the "production" of leisure.[20] I will now execute a partial and prelimi-

---

investment creation, and PPR execution, and, relatedly, (3) the percentages of the resources devoted to marginal uses of the above types that are withdrawn from the various other types of resource-uses. When my conclusions rely on my more-or-less-well-informed guesses about the magnitude of one or more relevant parameters, I will indicate that fact by referring to the guesses in question as "guestimates."

[20] Admittedly, a few relevant (natural) resources are also withdrawn from future uses (even if we classify the creation of the QV investment and execution of a production-

nary TBLE analysis of the aggregate distortion in the private cost of creating a marginal QV investment. The analysis will consider how Pareto imperfections inflate or deflate the private value to their alternative users of the resources that the creation of a marginal QV investment will withdraw respectively from the production of additional units of extant products, from the execution of production-process research, and from the "production" of leisure. I will then consider the aggregate distortion in the private cost of creating a QV investment that will be generated by the combined consequences of these three types of distortions.

1. The Aggregate Distortion in the Private Value of a Marginal Unit of Output ($\Sigma D(PV_{\Delta UO})$) and the Associated Distortion in the Private Cost of Creating a Marginal QV Investment

Because, in my judgment, a large portion of the economy's resources are allocated to unit-output-increasing uses, most of the resources used to create marginal QV investments are probably taken from the production of already extant products. For simplicity, I will assume that the creation of a marginal QV investment will reduce by one unit the output of each extant product whose unit output it affects.

I begin by analyzing the monopoly distortion in the private value of a marginal unit of output. At least if a unit-output producer does not engage in costless and perfect price discrimination, the imperfections in price competition he faces will almost always deflate the marginal revenue he obtains by selling his last unit of output.[21] For example, assume (1) that each block of resources that the creation of a marginal QV investment withdraws from a unit-output producer would have been used by that producer to produce an eleventh unit of output and (2) that this producer sells each of the ten units he produces without the resources in question for $10 but would have reduced the price on each to $9.50 if he had produced the eleventh unit because the eleventh-highest valuation of a unit of his product was $9.50.

---

process-research project as a contemporaneous use). I ignore this reality here both because the relevant resource-flow is almost certainly smaller than the flows analyzed in the text and because the sign of the aggregate distortion in the private value of such future uses is uncertain.

[21] On otherwise-Pareto-perfect assumptions, imperfections in price competition will not distort the private value of a marginal unit of output to an imperfect competitor who engages in costless perfect price discrimination because the marginal revenue such a producer will obtain by selling his last unit of output will equal its price, which will equal its allocative value (if the buyer of the relevant unit is a non-monopsonistic, sovereign maximizer whose consumption generates no externalities and whose purchase generates no sales taxes). The words "almost always" appear in the text because the profit-maximizing price of a non-discriminating imperfect competitor will also equal the marginal revenue his sale of his last unit of output generates if the demand curve he faces is horizontal (though above his marginal cost) at his profit-maximizing output.

On these assumptions, the monopoly distortion in the private value of each block of resources to its potential unit-output-increasing user and hence in its private cost to the marginal QV investor is (–$5). Thus, on these assumptions, the private value of these resources to the relevant unit-output producer—the marginal revenue he would obtain by selling the eleventh unit—is (11[$9.50]–10[$10]=$4.50), while the allocative value that the resource will produce in the relevant unit-output producer's employ, and hence the allocative cost of the QV investor's using these resources to create the marginal QV investment in question, is $9.50 (the dollar value to its consumer of the good the unit-output producer would have used them to produce).[22] Since the relevant private value is $4.50 and the relevant allocative value is $9.50, the monopoly distortion in the private value of the relevant marginal unit of output to its producer and hence in the private cost to the marginal QV investor of bidding these resources away from that producer is (–$5). Therefore, if the creation of a marginal QV investment withdraws N such blocks of resources from such unit-output producers, the monopoly distortion in the private cost of creating the marginal QV investment would equal (–$5)N.

Admittedly, an analysis of the relevant aggregate distortion would have to consider the distorting influences of any relevant non-sovereignty, non-maximization, monopsonies, sales taxes, externalities of production and consumption, and buyer surplus on the private value of the resources the marginal QV investor would use to create his QV investment to the unit-output producer who would place the second-highest private value on them. Nevertheless, I am confident that these imperfections would not cause the relevant aggregate distortion to have a different sign from the monopoly distortion. More specifically, I am confident that these imperfections would not change my conclusion that the private value of these resources to the potential unit-output-increasing user will be lower than the allocative value they would generate in his employ.

2. The Aggregate Distortion in the Private Value of a Marginal Production-Process-Research Project ($\Sigma D(PV_{\Delta PPR})$) and the Associated Distortion in the Private Cost of Creating a Marginal QV Investment

The creation of a marginal QV investment will also withdraw resources from contemporaneous investment in production-process research. As we have seen, the distortion in the private cost of these re-

---

[22] This latter conclusion reflects the fact that monopoly-distortion analyses proceed on the assumption that, aside from imperfections in price competition, the economy is Pareto-perfect. In an otherwise-Pareto-perfect economy, (1) the buyer in question is a non-monopsonistic, sovereign maximizer, (2) the buyer's purchase and consumption of the relevant unit generate no taxes, no buyer surplus, and no externalities, and (3) the production of the eleventh unit in question generates no externalities.

sources to the marginal QV investor will equal the distortion in the private value they would generate for the production-process researcher who bid the second-highest amount for them.[23] The private value of resources to a production-process researcher who is certain to make the discovery his PPR is designed to make sooner than anyone else has two components: (1) the gains the researcher obtains because his discovery reduces the costs that he (or a discovery-buyer or licensee) incurs to produce his pre-discovery output and (2) the gains the researcher obtains because, by reducing the marginal cost of production, the discovery enables him (or its other users) to profit by increasing his (or their) unit outputs. Monopoly deflates both these components of the private value of a PPR discovery.

Imperfections in price competition deflate the private gains that a production-process researcher can expect to obtain from the discovery-driven reduction in the cost of producing the relevant product's pre-discovery output, because such imperfections will deflate the pre-discovery private cost to him of the resources the discovery will enable him to save. Imperfections in price competition deflate the profits that the user of a marginal-cost-reducing discovery can obtain by expanding his output because such imperfections deflate the private value of the relevant units of output by more than they deflate their private cost. Each of these arguments requires some elucidation.

Imperfections in price competition deflate the private cost-savings a production-process discovery enables its user to secure on his pre-discovery output because (1) most of the resources the discovery will obviate its user's purchasing to produce his pre-discovery output will have been withdrawn from the unit-output production of imperfectly discriminating imperfect competitors and (2) as we have just seen, the private value of these resources to such alternative users would have been less than the allocative value their use of them would have generated. Thus, if we assume that the discovery will enable its user to produce his pre-discovery output with N fewer blocks of resources each of which will be allocated to a unit-output producer who will use it to produce an eleventh unit of output whose production will lead him to reduce his price from $10 to $9.50, the allocative value that the discovery will generate by reducing the cost its user must incur to produce his pre-discovery output will be $9.50N, while the private value it generates in this way for the discoverer will be $4.50N.

---

[23] The assumption that the marginal production-process-research firm from which the relevant resources would be withdrawn would always have been successful in being the first to make the production-process discovery the project was designed to yield is obviously unrealistic. The actual private value to a marginal production-process researcher of the resources he would use to execute his project equals (the value indicated in the text *times* the probability that he will make the discovery first) *minus* the risk costs he would have to incur to execute the project in question.

The conclusion that imperfections in price competition will also deflate the private gains a PPR discovery enables its user to obtain by increasing his unit output is based on two points. First, if the pre-discovery unit output of the relevant good is allocatively efficient, the private profits the discovery-user will realize by increasing his unit output post-discovery will be less than the allocative-efficiency gains the output-increase generates. This is true to the extent that the discovery leads its user to increase his output by more units than the number of units by which the associated output-increase reduces the unit outputs of each other individual good from whose production the discovery-user withdraws the resources he employs to produce the extra units of his own good. This conclusion reflects three facts: (1) the fact that the monopoly deflation in the private value to the discovery-user of the discovery-induced increase in output equals the average gap between the relevant seller's demand and marginal revenue curves between his pre-discovery and post-discovery outputs while the relevant monopoly deflation in the private cost that the discovery-user must incur to increase his output equals the gap between the demand and marginal revenue curves of the products that lose output when the discovery-user increases his output, (2) the fact that, *ceteris paribus*, the gap between price and marginal revenue increases with output, and (3) the fact that there is no reason to believe that the demand curve of the products to whose marginal production the marginal PPR project relates will be systematically and relevantly different from the weighted-average demand curve faced by the sellers whose outputs will drop when the discovery induces its users to expand their outputs.[24]

Second, the profits the discovery enables its user to realize by increasing his output will also tend to be deflated because, as this Article will demonstrate, the pre-discovery unit output of the goods to whose production the discovery relates tends to be sub-optimal from the perspective of economic efficiency (because ΣUO is sub-optimal from that perspective). Roughly speaking, this conclusion reflects the fact that the percentage by which the private cost of those resources a unit-output producer withdraws from alternative QV-investment-creating, PPR-executing, and leisure-producing uses is deflated, is lower on balance than the percentage by which the private cost of those resources a unit-output producer withdraws from alternative unit-output-producing uses is deflated (which, on the average, equals the percentage by which the private value of producing units of output is deflated by the monopoly or oligopoly power of the producer in question).[25] This conclusion is relevant because the profitability of producing units of output that were not

---

[24] For a diagrammatic exposition, see *Monopoly and Tort Law*, *supra* note 6, at 353–54.

[25] For a diagrammatic exposition, see *id.* at 361–62.

originally produced despite the fact that their production would have been allocatively efficient (because the relevant imperfections originally deflated the profitability of producing them) is always deflated.[26]

For the above reasons, I believe that the monopoly distortion in the private value of a marginal production-process-research project is negative. This implies that, *ceteris paribus*, the private cost to the QV investor of using the resources he withdraws from production-process research to create his investment is deflated.

Once more, of course, the assumption that the only type of relevant Pareto imperfection in the system is imperfections in price competition is inaccurate. Three additional types of Pareto imperfections have to be considered when analyzing the aggregate distortion in the private value of marginal production-process-research projects. First, to the extent that the production-process researcher is not the sole user of his discovery, the private value of his research-project will be deflated by any buyer surplus that buyers or licensees of his discovery will obtain from using his discovery. The cause of such surplus may be either the buyers' monopsony power or the inability of the seller to practice the price discrimination needed to eliminate such buyer surplus without incurring prohibitive private transaction costs.

Second, the private value of marginal PPR projects will also be deflated to the extent that the discoverer must increase the wages he pays his workers. These wages could rise (1) because regardless of whether the discovery-user's employees are unionized, the discovery-user is a monopsonist of labor and the discovery's use increases the workers' marginal-revenue-product curve or (2) because the monopoly power of the union that represents the discovery-user's employees enables them to secure a share of the gains that the discoverer would otherwise obtain from his discovery.

Third, the private value of marginal PPR projects will also be distorted by two different types of externalities. First, to the extent that the use of the production-process discovery reduces external costs of production, the private value of the associated production-process research will tend to be deflated, regardless of whether the discoverer uses his discovery himself. Second, the private value of a marginal PPR-project will also tend to be deflated to the extent that there is some possibility that the discovery it generates will impose costs on the discoverer by making him liable for any accident or pollution losses he generates because he failed to use the discovered process despite the fact that its use would reduce the certainty-equivalent accident or pollution losses he generates by more than it would increase other production costs. Production-process discoveries will increase the discoverer's legal liability if (1) he uses the discovery himself, (2) he is liable in tort and environmental law only if neg-

[26] For a diagrammatic exposition, see *id.* at 335–38.

ligent, (3) in practice, judges and juries do not find injurers negligent for failing to undertake production-process research, but (4) judges and juries find injurers negligent for failing to adopt production-processes they have discovered whose use would increase other production costs by less than it would reduce legally recognized accident-and-pollution costs.

The tendency of the discoverer's production-process research to change his legal liability is relevant because it involves a private loss that has no allocative counterpart. In fact, the private loss will probably be coupled with an allocative efficiency gain, since increased liability should induce the discoverer to engage in allocatively efficient avoidance.

The final "externality"[27] issue relates to the free-rider problems associated with knowledge-creation. Even in the presence of legal mechanisms that reward discoverers, production-process research may provide information to users who do not have to pay for it because (1) their use of it is not detected, (2) the discoverer is held not to have property rights in the knowledge, or (3) the users employ the knowledge in ways that are found not to violate the discoverer's property-rights (e.g., use it to develop alternative production-processes that will not be found to infringe the original discoverer's patents). More accurately, this issue will arise to the extent that the *ex ante* certainty-equivalent award the legal system offers the marginal production-process researcher falls below or exceeds the *ex ante* certainty-equivalent allocative value of his research-effort.

This "externality" issue is extraordinarily complex to analyze.[28] I will limit myself here to making only three observations on the topic. First, I want to outline the protocol the government should use to calculate the reward it should offer a production-process discoverer if it tries to increase economic efficiency by replacing the patent system with a system of direct government payments. If, for simplicity, we assume that all potential researchers are indifferent to risk, that is, are risk-neutral, one would calculate the necessary reward through the following two-step procedure. First, identify the potential researchers whose research would be part of the second-best most-allocatively-efficient production-process-

---

[27] The word "externality" is in quotes because the inflation in private value that would result from IP laws that offer a production-process researcher a certainty-equivalent reward that exceeds his research-effort's *ex ante* certainty-equivalent allocative value might better be classified as a negative externality distortion than as an externality distortion. I should add that discoverers may also be able to "internalize" what would otherwise be knowledge-creation-spillover externalities by taking advantage of the head-start their making the discovery gives them and/or by purchasing assets and businesses whose value will be increased by the discovery.

[28] I have completed two book manuscripts that devote considerable attention to this externality issue. Richard S. Markovits, First-Best, Second-Best and Third-Best Allocative-Efficiency Analysis: A Marine-Salvage Illustration (1992) (unpublished manuscript, available from the author); Richard S. Markovits, Antitrust and R&D Misallocation: A Partial Third-Best-Allocative-Efficiency Distortion and Policy Analysis (1996) (unpublished manuscript, available from the author).

research effort.[29] To do this, one would have to calculate the *ex ante* allocative benefits and allocative costs of each potential researcher's research-efforts, given the efforts of others.[30] Second, calculate the reward that should be offered to each participant in the second-best most-allocatively-efficient production-process-research effort by dividing the *ex ante* allocative value of that potential researcher's effort as part of the second-best most-allocatively-efficient research portfolio by the probability that he would make the discovery first if he participated in such a research-effort and subtracting from that sum the amount by which the relevant researcher's private PPR-costs were deflated. Even if the relevant PPR costs are not distorted, this optimal reward will not equal the allocative value of the discovery's being made when it was made as opposed to its never being made. Instead, even on this assumption, it will equal the allocative value of the relevant discovery's being made on the earlier date on which it would be predicted to be made if the discoverer's efforts were added to the other projects that belonged to the second-best most-allocatively-efficient research portfolio rather than on the later date it would be predicted to be made if the discoverer's efforts were withdrawn but all other members of that second-best most-allocatively-efficient set of projects were executed.[31]

The second point I wish to make about PPR-related externalities of knowledge-creation is that if, contrary to fact, under our current patent system a discoverer could obtain rewards for his discovery only by obtaining a patent, the patent system would tend to over-reward some discoverers and under-reward others. Put crudely, if all discoveries had equal allocative value, the current system would over-reward those researchers whose efforts advanced the expected date of the discovery they

---

[29] By "second-best most-allocatively-efficient production-process-research effort," I mean the effort that would be most allocatively efficient given the imperfections and misallocations in the system that affect the allocative value and cost of the production-process discovery.

[30] Since the allocative benefits that a particular production-process discovery will generate will depend not only on (1) the imperfections in the price competition that its users, its users' factor-rivals, and its users' product-rivals face, but also on (2) all those imperfections that affect the date on which complementary and substitute discoveries will be made, it will not be TBLE to calculate the *ex ante* allocative benefits of a particular PPR project without taking many, if not all, such imperfections into account. The allocative cost of a given PPR project equals the allocative value that the resources it would consume would generate in their actual alternative uses. Since that value will be affected if resources are misallocated to the alternative uses from which the PPR project's resources are withdrawn, it will usually be TBLE to take such misallocations and many if not all the Pareto imperfections that generated them into account when estimating the allocative cost of a PPR project.

[31] The calculation-procedure just described is also relevant in many other legal contexts. Examples include the determination of the percentage of the private value of rescued salvage it would be allocatively efficient to entitle a successful marine salvor to keep; the percentage of a found good it would be allocatively efficient to allow a finder to keep; and the percentage of the recovery in a class-action suit it would be allocatively efficient to award the plaintiffs' lawyers.

made by only a short period of time and would under-reward those whose research advanced the expected date of the discovery they made by fifty years.

The third point I wish to make at this juncture is that patent protection is less important than it may seem. Empirical research reveals that, because the information an applicant must make public to obtain a patent enables others to invent around his patent, firms in many American industries find that secrecy, trade-secret protection, and lead-time advantages are preferable to patents.[32]

I have no idea whether our current system of rewarding researchers over-rewards or under-rewards marginal production-process-research investors. To analyze this issue, one would have to gain some sense of (1) the profits that various subsets of successful marginal production-process researchers can make if others can copy their discoveries as soon as they use them by buying up assets that are complementary to the activities their discoveries will promote (assets whose private value will be increased by the discoveries), (2) the average breadth of the patent, copyright, and trade-secret protection given to successful production-process researchers in each such subset, (3) the profits that discoverers in each such subset can realize by keeping the details of their discoveries secret, and (4) the *ex ante* allocative value that the marginal production-process researcher in each such subset generates by advancing the expected date on which the discovery he is trying to make will be made. I have insufficient hard data on and insufficient experience with these parameters to be willing to venture even a guess about whether across all cases marginal production-process researchers would tend to be over-rewarded or under-rewarded by our current system from the perspective of allocative efficiency if imperfections in seller and buyer competition, union monopolies, and conventional externalities of production could be ignored. However, I do suspect that taken together these latter imperfections deflate the private value of the average marginal production-process-research project sufficiently to justify the conclusion that, even when the distorting impact that our IP law would have in the absence of such imperfections is taken into account, the aggregate distortion in the private value of marginal production-process-research projects is negative. This conclusion implies that the private cost to marginal QV investors of the resources they bid away from production-process researchers is deflated.

---

[32] Richard C. Levin *et al.*, *Appropriating the Returns from Industrial Research and Development*, *in* 3 BROOKINGS PAPERS ON ECONOMIC ACTIVITY 783 (1987). Since, *ceteris paribus*, allocative efficiency will be increased if discoveries are used by everyone to whom they have value (indeed, that is why a government-reward system may be more attractive than a patent-copyright system), a full analysis of the most allocatively efficient reward to give a discoverer would have to consider the effects of the alternative option of keeping discoveries secret and not applying for the available reward.

3. The Aggregate Distortion in the Private Value of a Marginal Unit of Leisure ($\Sigma D(PV_{\Delta Lei})$ and the Associated Distortion in the Private Cost of a Marginal QV Investment

QV investors withdraw some of the resources they use to create their QV investments from the "production of leisure." As the text has shown, this implies that any distortion in the private value of leisure to its producer/consumer will convert into an identical distortion in the private cost of creating a marginal QV investment.

In my judgment, four types of Pareto imperfections create significant distortions in the private value of leisure to its producer/consumers: imperfections in competition, non-sovereignty, externalities, and taxes on the margin of income. I will now analyze the distorting impact of each of these types of Pareto imperfections on the private value of leisure to its producer/consumers and hence on the private cost to a QV investor of creating a marginal QV investment with resources that would otherwise have been devoted to the "production" of leisure.

Imperfections in competition distort the private value of leisure to the potential workers who forego leisure-consumption to create a marginal QV investment by distorting the private cost to them of the complements they consume together with their leisure. More specifically, because I believe that the price of goods virtually always exceeds not only their marginal private cost but also their marginal allocative cost, I believe that imperfections in seller competition tend to deflate the private value of leisure to its consumers by inflating the private cost to them of complements of leisure. Admittedly, the correlative deflation in the private cost of creating a marginal QV investment will be somewhat offset by the tendency of imperfections in competition to inflate the private cost of labor to the QV investor by inflating the private cost to his potential workers of the goods they use to work for him, such as work clothes, automobiles used to commute, and gasoline used to commute. However, at least in part because the relevant workers also use some of these latter goods when consuming leisure, I would guess that on balance the preceding considerations favor the conclusion that imperfections in competition deflate the private cost to a marginal QV investor of those resources he withdraws from the production of leisure to create his QV investment.

The private cost to a marginal QV investor of withdrawing resources from the production of leisure is also distorted by the non-sovereignty of the potential workers in question, i.e., by their misestimating the value of leisure to them. Admittedly, I have no particular expertise on this issue, but my guess is that members of our culture tend to undervalue leisure. If so, this mistake will tend to deflate the private cost to a QV investor of the resources he bids away from leisure-production to create his QV investment.

The third type of Pareto imperfection that is relevant in the current context is externalities. Two types of externalities are relevant in this context—the conventional pollution and congestion externalities that leisure-consumption and labor generate and what I will call the "interpersonal externalities" it generates. Pollution and congestion externalities will inflate the private value of leisure and hence the private cost the QV investor must incur to bid resources away from the production of leisure to the extent that (1) the externalities the QV investor prevents by inducing workers to forgo the consumption of leisure (because they would have driven around, used crowded tennis courts or golf courses, or visited crowded shopping malls in their leisure time) exceeds (2) the extra commuting pollution and congestion costs they generated when working for the QV investor. I suspect that whether such externalities inflate or deflate the private cost of a marginal QV investor will be substantially affected by whether he reduces leisure by (1) increasing the number of hours that individual workers who would have worked on the days in question in any event perform each day they work or (2) increasing the number of "days" of labor performed (that is, the number of commutes). The interpersonal externalities that might be generated by the consumption of the relevant leisure could also cut in either direction. Thus, to the extent that the relevant leisure would have been consumed by parents who would have used it to play constructively with their kids or coach Little League, its consumption would have generated external benefits, its private value to the consumer would be deflated, and the private cost the QV investor would have to incur to withdraw resources from its production would be deflated. In the other direction, to the extent that the consumer of the relevant leisure would have spent the time being a nuisance around the house or berating his children non-constructively, its consumption would have generated external costs, its private value to its consumer would be inflated, and the private cost to the QV investor of withdrawing resources from its production would also be inflated.

The fourth and final type of Pareto imperfection that is relevant in the current context is taxes on the margin of income. To the extent that the relevant workers' decisions to allocate their time to creating a marginal QV investment rather than to producing/consuming leisure decreases the transfer payments (e.g., unemployment insurance payments or welfare benefits) they receive or increases the taxes they pay, the private cost of hiring them will be inflated. Given the height of the applicable tax rates, I suspect that this tax-and-transfer-related inflation of the private cost a QV investor will have to incur to bid resources away from leisure-production to create his QV investment may be sufficiently big to cause the aggregate distortion in the relevant private cost to be positive.

* * *

I have now concluded that the relevant imperfections (1) deflate the private cost the marginal QV investor must incur to create his marginal QV investment by withdrawing resources from (A) contemporaneous marginal unit-output production and (B) contemporaneous marginal PPR-execution and (2) inflate the private cost to the QV investor of resources that he withdraws from leisure-production to create his QV investment. Although, obviously, the following judgment reflects a wide range of other empirical assumptions as well, my assumption that the vast majority of the resources used to create QV investments are withdrawn from unit-output production leads me to conclude that $\Sigma D(PC_{\Delta QV})$ is almost certainly substantially negative.

B. The Aggregate Distortion in the Private Value of a Marginal QV Investment

The private value of a QV investment equals the operating profits the QV investor can realize by using his QV investment, that is, by producing and selling actual units of his new product-variant instead of abandoning the relevant design or by operating rather than costlessly destroying his new distributive outlet or capacity. This analysis of the aggregate distortion in the private value of a marginal QV investment contains three parts. The first focuses on the distortion in the private value of a marginal QV investment generated by the imperfections in the price competition faced by the sellers whose outputs will be reduced when a marginal QV investment is used. The second focuses on the distorting influence of other imperfections in seller competition, *viz.*, imperfections in QV-investment competition,[33] imperfections in the price competition that the marginal QV investor faces when selling his newly designed product-variant, operating his new outlet, or using his added capacity or inventory; and the monopoly power of the marginal QV investor's unionized employees. Finally, the third part focuses on the distorting impact

---

[33] In my terminology, "quality-or-variety-increasing-investment competition" refers to the process by which the introduction of additional QV investments into a given arbitrarily designated portion of product-space ("ARDEPPS") eliminates the supernormal profits yielded by QV investments. I have substituted the concept of ARDEPPS for the standard term "market" because in our actual imperfectly competitive world, markets cannot be defined non-arbitrarily. For a detailed demonstration of this conclusion, see Richard S. Markovits, *On the Inevitable Arbitrariness of Market Definitions and Inevitable Cost-Ineffectiveness of Market-Oriented Approaches to Competitive-Impact Prediction*, ANTITRUST BULL. (forthcoming 2002). The intensity of QV-investment competition can be measured either by the equilibrium supernormal profit-rate yielded by an ARDEPPS' most profitable QV investments or by the difference between its equilibrium QV-investment level and the level of ARDEPPS QV investment that would cause its most profitable QV investments to yield just normal rates of return. For a further elaboration of these concepts, see Richard S. Markovits, *Predicting the Competitive Impact of Horizontal Mergers in Our Monopolistically Competitive World: A Non-Market-Oriented Proposal and Critique of the Market Definition-Market Share-Market Concentration Approach*, 56 TEX. L. REV. 587, 660–73 (1978).

of other types of Pareto imperfections on the private value of a marginal quality-or-variety-increasing investment.

1. The Distortion in the Private Value of a Marginal QV Investment Generated by the Imperfections in the Price Competition Faced by the Producers Whose Unit Outputs Are Reduced by the Use of the Marginal QV Investment

The resources that the marginal QV investor will employ to use his QV investment will be withdrawn from a variety of different sources: from the production of leisure, from the execution of marginal production-process-research projects, and from the production of marginal and intra-marginal units of existing products. Because the sale of the new product or service will reduce the sales of its product-rivals' products directly, I assume that the overwhelming majority of the resources the marginal QV investor employs to use his QV investment will in effect be withdrawn from unit-output-increasing uses by the producers of the products that compete with the new product or service. I will therefore confine the analysis of the price-competition-imperfection distortion in the private value of a marginal QV investment to the distortion that the imperfections in the price competition faced by the prospective product-rivals of the product-service the marginal QV investment will create generate in the private value of that marginal QV investment.

Imperfections in the price competition faced by the marginal QV investor's product-rivals distort the private value of the marginal QV investment by deflating the private value to these product-rivals of the resources that the marginal QV investor bids away from them when he uses his QV investment. More specifically, to the extent that the QV investor's production of the new product withdraws resources from the production of old products produced by imperfect competitors for whom marginal revenue is less than price, the imperfectness of the price competition faced by these firms will inflate the private value of the marginal QV investment by deflating the private cost to the QV investor of using his QV investment, i.e., the private cost to him of the resources he combines with his QV investment to produce actual units of his new product or operate his new distributive outlet or capacity.

For example, assume that the marginal QV investor will sell one unit[34] of his new product for $100 and that the private cost to him of producing that unit is $45. If the resources he used to produce that unit were withdrawn from the production of the eleventh unit of ten other goods by sellers who would have reduced their prices from $10 to $9.50 to sell the eleventh units in question, i.e., by 10 sellers each of whom would have

[34] This assumption is made to eliminate the possibility of a buyer-surplus deflation in the private value of a marginal QV investment, which will be analyzed later in this Section.

obtained marginal revenue of $4.50 by selling the eleventh units in question for $9.50, the allocative cost of the marginal QV investor's using his marginal QV investment would be 10($9.50) or $95 (on our current otherwise-Pareto-perfect assumption) and not the infinitesimally-higher-than-$45 sum he had to pay to obtain the resources in question. In this case, although the private value of the marginal QV investment will be ($100-$45)=$55, the allocative value of the marginal QV investment will be ($100-$95)=$5. The relevant imperfection in price competition will have inflated the private value of the investment ($PV_{\Delta QV}$) by $50 by deflating by $50 the private cost of using the relevant marginal QV investment.

2. The Distortion in the Private Value of a Marginal QV Investment Generated by (1) the Imperfections in QV-Investment Competition Facing the Marginal QV Investor, (2) the Imperfections in Price Competition the Marginal QV Investor Himself Faces When Selling His New Product or Service, and (3) the Monopoly Power of the Marginal QV Investor's Unionized Employees

When QV-investment competition is imperfect in any ARDEPPS, a QV investment will sometimes reduce the profits that the investor's intra-marginal projects yield by taking sales away from them directly, by inducing rivals to make non-retaliatory non-QV-investment responses (e.g., pricing or advertising moves), or by inducing rivals to make non-retaliatory QV investments that they would not otherwise have found profitable. At other times, a marginal QV investor who faces imperfect QV-investment competition will find that his marginal QV investment increases the profits his intra-marginal QV investments yield by deterring QV investments by competitors that would have reduced his pre-existing projects' profit-yields in the first two ways just listed to a greater extent than his marginal QV investment reduced their profit-yields relative to the *status quo ante*. Both the private losses and the private gains just described[35] are "distorting" in that no equal allocative losses or gains are associated with them. I believe that the private value of more marginal QV investments is inflated than deflated by these imperfect-competition-related distortions. However, imperfections in competition will deflate the private value of some marginal QV investments on this account.

The second imperfection in competition that is relevant in the current context is the most important to be considered in this section: the imperfectness of the price competition faced by the marginal QV investor

[35] I denominate the relevant private losses "monopolistic QV investment disincentives" or "natural oligopolistic QV investment disincentives" (in the case in which the QV investment induces a rival to make a QV investment the latter would not otherwise have made). I denominate the relevant private gains "monopolistic QV-investment incentives."

himself, which causes the buyer-surplus deflation in the private value of the marginal QV investment. Virtually all marginal QV investors' sales of their new product-variants or operation of their new distributive outlets or capacity will generate buyer surplus for the buyers of the products or services the marginal QV investments created.[36] To the extent that this is the case, the private value of the marginal QV investment will be *ceteris paribus* deflated inasmuch as buyer surplus represents an allocative benefit generated by the marginal QV investment's use that is external to the marginal QV investor.

However, although this buyer-surplus deflation in the private value of a marginal QV investment is substantial, it is almost certainly far smaller than the inflation in the private value of using a marginal QV investment that is caused by the deflation in the private cost of using that investment generated by the imperfections in the price competition faced by the marginal QV investor's product-rivals and other factor-rivals. To see why, two relationships should be recalled. First, the per-unit-of-output buyer-surplus deflation in the private value of a marginal QV investment equals the average distance between the new product's demand and marginal revenue curves between the vertical axis and the new product or service's actual output, while the per-unit-of-output deflation-in-resource-cost inflation in the private value of a marginal QV investment equals the weighted-average gap between the price and marginal revenue curves of the marginal QV investment's product-rivals at the outputs they produced prior to the marginal QV investment's being used.[37] And second, the distance between a seller's demand and marginal revenue curves increases with his output. An example may illustrate the force of these considerations: if (1) all the relevant demand curves are linear, (2) none of the relevant sellers practices price discrimination, (3) the demand and marginal cost curves for the product created by the marginal QV investment equal (roughly speaking) their weighted-average counterparts for the old products whose unit outputs will be reduced when the marginal QV investment is used, and (4) the use of the QV investment reduces by one unit the outputs of all those old products whose outputs it reduces, the buyer-surplus deflation in the private value of a marginal QV investment will equal one-half the resource-cost-deflation-related inflation in the private value of a marginal QV investment since the average distance between a linear demand curve and its associated conventional marginal revenue curve between the vertical axis and its actual output will be one-half the distance between these two curves at its actual output. Although

---

[36] This buyer surplus would be over and above any surplus these buyers would have realized prior to the marginal QV investment's execution on the purchases they forewent to purchase units of the good or service the marginal QV investment created.

[37] This supposes that the output of each product-rival who lost sales when the marginal quality-or-variety-increasing investment was used was reduced by one unit when the marginal quality-or-variety-increasing investment was used.

the fact that, in reality, the vast majority of the individual old products that lose output when the marginal QV investment is used will lose more than one unit of output implies that the ratio of the deflation-in-resource-cost-related inflation in the private value of the marginal QV investment to the buyer-surplus deflation in that private value will be somewhat below two, I would not expect this ratio to be much below two.

The third imperfection in competition that is relevant in the current context is the monopoly power of the unionized employees of the marginal QV investor. To the extent that the QV investor's employees can secure a share of the operating profits the use of the marginal QV investment would otherwise yield, its private value will be deflated.

Although I admit that the net effect of all three of these additional imperfections in seller competition will be to deflate the private value of a marginal QV investment by a considerable amount, I am confident that the net deflation in the private value they generate will be far smaller than the inflation in the private value generated by the imperfections in the price competition faced by the unit-output producers from whom the marginal QV investor withdraws the resources he employs to use his QV investment.

3. The Distortion in the Private Value of a Marginal QV Investment Generated by Pareto Imperfections Other Than Imperfections in Seller Competition

Three other types of Pareto imperfections may significantly distort the private value of a significant number of marginal QV investments: monopsony, taxes on the margin of income, and Galbraithian and knowledge-creation externalities.

The monopsony that is relevant in the current context is the monopsony power of the buyers of the right to use any QV investment that is not used exclusively by its creator, e.g., independent buyers of the right to produce the technologically or commercially innovative product that a particular marginal QV investment created. Although I am certain that the overwhelming majority of QV investments are not innovative and am reasonably confident that the vast majority of product-innovations are produced by the firms that developed them, bilateral monopoly may sometimes deflate the private value of marginal QV investments to creators who sell the right to produce their product to one or more independents.

The tax imperfections that are relevant in the current context are accelerated-depreciation provisions, provisions that allow research-expenditures that do not create physical capital to be expensed immediately even though they have a long pay-off period, investment and R&D tax-credit provisions, and provisions that encourage investment in such places as Puerto Rico. In my judgment, these provisions often critically inflate the

private value and profitability of marginal QV investments—indeed, they will often cause QV investments to be profitable that would not otherwise be profitable.[38] This conclusion reflects my belief that these provisions are more valuable to QV investors (and to production-process researchers) than to producers of products already in production. They will therefore have a tendency to render profitable some QV investments that would otherwise be unprofitable even after one considers the consequences of their critically affecting the profitability of some investments (say, in plant modernization) that benefit unit-output production. Admittedly, it is somewhat artificial to consider these tax distortions in a section concerned with the aggregate distortion in the private value of marginal QV investments. It would be more accurate simply to indicate that they will tend to inflate critically the profitability of some QV investments than to assign their distorting effect to either the aggregate distortion in the private value of marginal QV investments or to the aggregate distortion in their private cost.

The first type of externality distortion in the private value of a marginal QV investment is generated by what might be termed "Galbraithian external costs." A QV investment will generate such costs if it creates a new product whose introduction reduces the value of pre-existing products to those of their owners who value having "the latest thing." Since these external costs are not internalized to the QV investor who has introduced the new product-variant, they inflate the private value of marginal QV investments to those who make them.

The second type of "externality" distortion in the private value of a marginal QV investment is the distortion in the private value of technologically-or-commercially-innovative marginal QV investments that arises when the combination of patent, copyright, and trade-secret law and the ability of discoverers to secure profits by taking advantage of their head-start and/or by buying assets or businesses whose value their discovery will enhance does not equate (1) the *ex ante* certainty-equivalent profits expected to be yielded by the relevant product-innovation research-efforts with (2) the amount by which these research-efforts would increase allocative efficiency if the only Pareto imperfections in the system were knowledge-creation spillovers and the tendency of any given research-project to reduce the probability that other researchers will discover first whatever the research-project in question is designed to discover. Unfortunately, the sign and value of this distortion are equally

---

[38] For an explanation and illustration of the way in which such tax provisions can make otherwise unprofitable investments profitable, see Calvin H. Johnson, *Tax Shelter Gain: The Mismatch of Debt and Supply Side Depreciation*, 61 TEX. L. REV. 1013 (1983) [hereinafter *Tax Shelter Gain*]. Admittedly, Johnson's analysis focuses on some provisions that have been removed and others whose force has been reduced. Thus, the investment tax-credit created by § 46 of the Internal Revenue Code of 1954 was basically repealed by § 211(a) of the Tax Reform Act of 1986, Pub. L. No. 99-514, 100 Stat. 2085.

as hard to predict as are their PPR counterparts. In fact, I am equally uncertain about them as I was about their respective PPR counterparts. Fortunately, however, my ignorance is less troublesome in the current context because the percentage of QV investments that are innovative in the sense that is relevant in this context is probably extremely low.

All told, I am confident that the imperfections discussed in this subsection and the various other non-monopoly imperfections that could distort the private value of marginal QV investments on their own will not fully offset (if they offset at all) the positive distortion in that private value caused by the imperfections in the price competition faced by the unit-output producers from whom the use of the marginal QV investment withdraws resources. Indeed, I am certain that the aggregate percentage-distortion in that private value is not only positive but substantial.

C. The Aggregate Distortion in the Profits Yielded by Marginal QV Investments

The preceding two sections respectively argued that, on balance, the relevant Pareto imperfections deflate the private cost of marginal QV investments and inflate their private value. Obviously, these two conclusions imply that the profits of marginal QV investments are inflated—that their profitability exceeds their allocative efficiency—and that as a consequence our economy generates total-QV-investment ("ΣQV") misallocation by allocating too many resources to QV investment.

Two points need to be made about this conclusion. The first is that since the QV investments on which this analysis has focused are marginal in the sense of being the last QV investments to be made in their respective ARDEPPSes but not in the sense of being minute, that is, since QV investments are lumpy or incremental, the conclusion that the aggregate distortion in the profitability of marginal QV investments is positive in a given ARDEPPS does not guarantee that the marginal QV investment in that ARDEPPS is allocatively inefficient. Admittedly, inasmuch as the lumpiness of QV investments makes it possible for the profits yielded by a marginal QV investment to be not only positive but higher than the aggregate distortion in those profits, the allocative efficiency of the investment in question may be positive even when the aggregate distortion in its profits is positive. However, in my judgement, marginal QV investments in virtually all ARDEPPSes are allocatively inefficient, i.e., the percentage-inflation of the profits of marginal QV investments is virtually always very high, much higher than the positive rate of return yielded by the relevant marginal QV investments.

The second point I wish to emphasize at this juncture is that the conclusion that the profits of the marginal QV investment in an ARDEPPS are inflated does not imply that all QV investments in that ARDEPPS are allocatively inefficient. At most, it implies that usually the marginal QV

investment in that ARDEPPS and one or more intra-marginal QV investments in that ARDEPPS (more specifically, all QV investments in that ARDEPPS whose profits are lower than the aggregate distortion in their profits) are allocatively inefficient. The profits of many intra-marginal QV investments in each ARDEPPS, however, will be higher than the distortion in their profits. For example, if there are 43 QV investments in some ARDEPPS whose marginal QV investment's profits are inflated—that is, in which the aggregate distortion in the profits of its marginal QV investment is positive—my argument might imply that it would be allocatively efficient for there to be 38 or 34 QV investments in that ARDEPPS, not that it would be allocatively efficient for there to be just one QV investment in that ARDEPPS.

III. The Aggregate Distortion in the Profits Yielded by Marginal Production-Process-Research Projects: A Partial and Preliminary TBLE Analysis of the Causes of Total Production-Process Research (ΣPPR) Misallocation

Part III executes a partial and preliminary TBLE analysis of the aggregate distortion in the profits yielded by marginal PPR projects that parallels the preceding analysis of the aggregate distortion in the profits yielded by marginal QV investments. Since virtually all the relevant intermediate results have been established in the course of analyzing the Pareto imperfections that cause total QV investment to be misallocatively high, Part III is quite short.

A. The Aggregate Distortion in the Private Cost of a Marginal Production-Process-Research Project

The execution of marginal PPR projects that increase the total amount of resources devoted to PPR in the economy will withdraw resources primarily from the production of marginal units of output, the creation of marginal QV investments, and the production of leisure. As I have demonstrated, the relevant imperfections will deflate the private value of marginal units of output to their producers, inflate the private value of marginal QV investments to those who create them, and inflate the private cost of attracting resources away from the production of leisure.

These results imply that the sign of the aggregate distortion in the private cost of marginal PPR projects will depend on the following two sets of factors: (1) the percentages of the resources used to execute marginal PPR projects that are withdrawn from contemporaneous unit-output production, QV-investment creation, and leisure production and (2) the absolute percentage-distortions in the private value of the resources used

to produce marginal units of output, to create marginal QV investments, and to "produce" marginal units of leisure.

In my judgment, on balance, the relevant imperfections deflate the private cost of executing marginal PPR projects. I reach this conclusion despite my belief that the percentage-inflation in the private value of a marginal QV investment is much higher than the weighted-average percentage-deflation in the private value of marginal units of output and marginal units of leisure. More specifically, I reach this conclusion for two reasons: (1) I assume that producers withdraw a very low percentage of the resources used to execute marginal PPR projects from the production of leisure and(2) I suspect that since, in my opinion, at least five or six times more resources are devoted to unit-output production than to QV-investment creation in our economy and the percentage of all QV investments that involve technological research that employs the type of technically skilled personnel who will also be used to execute PPR projects is low, the execution of PPR projects will not withdraw significantly more than one-fifth as many resources from QV-investment-creating uses as from unit-output-increasing uses.

I should note that the argument that the aggregate distortion in the private cost of marginal PPR projects is negative is an argument against interest, i.e., this argument cuts against my ultimate conclusion that the aggregate distortion in the profits yielded by marginal PPR projects is negative and that as a result we currently allocate too few resources to PPR from the perspective of allocative efficiency.

B. The Aggregate Distortion in the Private Value of Marginal Production-Process-Research Projects

The preceding analysis of the Pareto imperfections that are relevant to my conclusion that we currently allocate more resources to QV-investment creation than is allocatively efficient explained why the relevant imperfections in both seller and buyer competition seem likely to deflate substantially the private value of marginal PPR projects to their owners. It also explained why, on the average, this average monopoly and monopsony deflation in that private value is unlikely to be fully offset by the failure of our innovation law to internalize accurately the *ex ante* certainty-equivalent allocative value of marginal PPR projects. Even if in an otherwise-Pareto-perfect economy our current knowledge-creation law would overinternalize[39] the allocative benefits yielded by the average QV investment that is truly innovative, the percentage of QV investments that generates discoveries that are eligible for innovation-law protection is low. Because this percentage is low, the associated externality inflation in

---

[39] By "overinternalize" I mean "cause the certainty-equivalent private value of the relevant QV investments to exceed their *ex ante* allocative value."

the private value of marginal PPR projects is almost certainly smaller than the sum of the average monopoly and monopsony deflations in the relevant private value. Therefore, I believe not only that the average aggregate distortion in the private value of marginal PPR projects is negative but also that the aggregate distortion in the private value of virtually all individual marginal PPR projects is negative.

C. The Aggregate Distortion in the Profits Yielded by Marginal Production-Process-Research Projects

The preceding two sections concluded, *inter alia*, that the relevant imperfections deflated the private cost of executing marginal PPR projects, deflated the private value of the average marginal PPR project, and deflated the private value of virtually all individual marginal PPR projects in the economy. These conclusions imply that the sign of the aggregate distortion in the profits yielded by marginal PPR projects depends on whether the deflation in the private value of virtually all marginal PPR projects exceeds the deflation in their private cost. In my judgment, it does.

This conclusion primarily reflects three premises: (1) Marginal PPR projects generate PPR discoveries that save more resources than the execution of these projects withdraws from unit-output-increasing uses. This premise is relevant because it implies that imperfections in price competition deflate the private value of PPR discoveries by more than they deflate the private cost of production-process research; (2) the above deflation in the profits yielded by marginal PPR projects is virtually never fully offset by any inflation in those profits that our patent, copyright, and trade-secret laws would generate in its absence;[40] and (3) the profits yielded by marginal PPR projects are also deflated by other imperfections—most importantly, by (A) the monopoly power of the production-process researcher's unionized employees, (B) the monopsony power (in the discovery-buying context) of the independent paying-users of the PPR discovery, (C) the traditional accident and pollution externalities that the relevant PPR discovery's use will prevent, and (D) the tendency of the relevant PPR discovery to increase the tort-law and environmental-law liability of those marginal PPR discoverers who are in a position to use their own discoveries and who are liable only if they have acted "negligently" as that concept has been applied in the relevant legal system (i.e., are liable for failing to adopt a known (discovered) production-technique whose use would reduce accident-and-pollution costs by

---

[40] It is not clear that those laws would raise the certainty-equivalent profits expected to be yielded by marginal PPR projects above their *ex ante* allocative value in an otherwise-Pareto-perfect world. Even if they would, any such profit-inflation would be extremely unlikely to cancel out the profit-deflation described above.

more than it would increase other private costs of production but are not liable for failing to do PPR whose *ex ante* private cost is lower than the amount by which it would be predicted to reduce the sum of private production costs and accident and pollution costs).[41]

The conclusion that the aggregate distortion in the profits yielded by marginal PPR projects is negative implies that our economy tends to generate ΣPPR misallocation by allocating too few resources to PPR from the perspective of allocative efficiency.

IV. The Aggregate Distortion in the Profits Yielded by Producing Marginal Units of Output: A Partial and Preliminary TBLE Analysis of the Causes of Total-Unit-output (ΣUO) Misallocation

Part II of this Article analyzed the aggregate distortion in the profits yielded by marginal QV investments on the counterfactual assumption that their creation and use would not withdraw any resources from the creation of alternative QV investments, that is, on the assumption that their creation would increase the amount of resources our economy devotes to QV-investment creation, because this is the assumption that is relevant for analyzing total-QV-investment (ΣQV) misallocation. Because this assumption is appropriate for the analysis of our tendency to allocate too few resources to the production of existing products (to cause total-unit-output misallocation), Part IV analyzes the aggregate distortion in the profits yielded by the production of marginal units of output on the assumption that the production of the units in question will not withdraw resources from the production of additional units of other products, i.e., on the assumption that their production will increase the amount of resources our economy devotes to unit-output production. Thus, just as Part II ignored QV-investment-to-QV-investment resource-flows and inter-ARDEPPS and intra-ARDEPPS QV-investment misallocation, this sec-

---

[41] The text has focused on the *aggregate* distortion in the private profits yielded by different individual marginal PPR projects and the average value of this aggregate distortion across all marginal PPR projects. If the policy question is the allocative efficiency of increasing PPR in all ARDEPPSes, this average aggregate distortion is highly relevant. If, however, the policy question is the possible allocative efficiency of intervening more selectively—say, to make the breadth of patent-coverage vary with the *ex ante* certainty-equivalent allocative value of particular successful marginal PPR projects, information on the average distortion in the profits yielded by marginal PPR projects is less relevant than estimates of the distortion in the profits yielded by particular marginal PPR projects. Even if the average distortion is zero, a good deal of PPR misallocation will be present if $\Sigma D(P\pi_{\Delta PPR})$ is positive in some ARDEPPSes and negative in others since, from the perspective of allocative efficiency, too much PPR will be executed in the former ARDEPPSes and not enough in the latter. In this context, two wrongs do not make a right—that is, they do not offset each other even if they cut in opposite directions. Even if the average year-round temperature in some city were 68°F, that would not imply that the city has a pleasant climate since it might have an average summer temperature of 104°F and an average winter temperature of 14°F.

tion will ignore unit-output to unit-output resource-flows and the kind of relative-unit-output ("RUO") misallocation on which top-level allocative-efficiency analyses[42] have traditionally concentrated. On this basis, in Part IV I analyze the aggregate distortions in the private cost, private value, and profits yielded by the production of marginal units of output.

A. The Aggregate Distortion in the Private Cost of Producing Marginal Units of Output

We are currently concerned with the distortion the relevant imperfections generate in the private cost to a producer of the resources his production of a marginal unit of output withdraws from leisure-production, the execution of marginal PPR projects, and the creation of marginal QV investments. Part II's conclusions imply that the relevant imperfections will (1) deflate the private cost to these actors of the resources their production of marginal units of old products withdraws from the execution of marginal PPR projects (because the imperfections in question deflate the private value of marginal PPR projects), (2) inflate the private cost to producers of marginal units of old products of the resources their production withdraws from the creation of marginal QV investments (because the imperfections in question inflate the private value of marginal QV investments), and (3) inflate the private cost to these actors of the resources their production withdraws from the production of leisure. To be frank, my guesses about the percentage of the resources that marginal unit-output production consumes that are withdrawn from these three sources are too uncertain for me to be willing to predict the sign of the aggregate distortion in the private cost of marginal units of output.

B. The Aggregate Distortion in the Private Value of a Marginal Unit of Output to Its Producer

However, I am confident not only that the monopoly distortion in the private value to a producer of his marginal unit of output is negative (because marginal revenue is less than price) but that, on the average, the absolute value of this monopoly distortion is higher than any plausible average inflation in the relevant private value attributable to the other imperfections that can distort that value—most importantly, the net external costs that the producer of a marginal unit of output generates when

---

[42] Top-level misallocation is present to the extent that a transaction-costless shift in the set of product-types produced (including leisure, work of different degrees of attractiveness, future as opposed to present goods, the total amount of QV investment, and the particular QV investments that are made) or a change in the proportions in which goods in production are produced would give its beneficiaries a total equivalent-dollar gain that would exceed the total equivalent-dollar loss it would impose on its victims.

producing it and that the buyer of a marginal unit generates when shopping for and consuming it. Indeed, even if, as I suspect, such externalities do bias resource allocation toward unit-output production from the perspective of allocative efficiency (because the external costs of the production and consumption of marginal units of extant products constitute a higher percentage of their private marginal costs than the external costs of production and consumption generated by the creation and use of marginal QV investments or the execution and use of marginal PPR projects constitute of their respective total cost), this inflation will be at least somewhat offset by the tendency of tax laws to favor QV investment and PPR over unit-output production. Tax laws favor QV and PPR investment because those laws' investment-credit, accelerated-depreciation, and "expensing of capital expenditure" provisions are less valuable to unit-output producers than to QV investors or owners of PPR projects, given that a higher percentage of QV-investment-related and PPR expenses consists of "investment-expenditures."

C. The Aggregate Distortion in the Profits Yielded by the Production of a Marginal Unit of Output

Although I admit to uncertainty about the sign of the aggregate distortion in the private cost of producing a marginal unit of output, I am confident that the relevant imperfections deflate the profits yielded by the production of a marginal unit of output. Largely because of the effects of imperfect price competition, the deflation in the private value of a marginal unit of output to its producer will be higher than any plausible deflation in the private cost he has to incur to produce it. This conclusion reflects the fact that, while the private cost of the resources that marginal unit-output production withdraws from the production of leisure and from the creation of QV investments will be inflated by taxes on the margin of income and imperfections in price competition, nothing similar will inflate the private value the unit-output producer obtains from his production. The conclusion that the aggregate distortion in the profits yielded by the production of a marginal unit of output is negative implies that our economy generates total-unit-output (ΣUO) misallocation by allocating too few resources *in toto* to the production of unit output in the sense that allocative efficiency would be increased if more physical units of extant products were produced and fewer resources were allocated to the combination of QV-investment creation, PPR execution, and leisure production.

V. Policy Implications

Let me preface this Part with three caveats. The first is an expertise caveat (an assessment of my competence). I have real expertise in anti-

trust: I know a great deal not only about how the courts have interpreted and applied the American statutes and how these statutes should be interpreted and applied but also about the business behavior to which antitrust laws relate and the extent of monopoly and monopsony imperfections. I am not an expert in American tax law and do not have any special insight into the practicability of particular tax approaches, though I have studied substantive tax law and my welfare-economics training does give me some public-finance expertise. The previous sentence applies as well to my expertise in American tort law. I know very little about patent, copyright, and trade-secret law—either about the law on the books or about the way in which the law has actually been applied—though, once more, my welfare-economics training is relevant to knowledge-creation policy.

The second caveat is a dearth-of-empirical-information caveat. Although a few of the results on which the following analyses rely can be established on a purely *a priori* basis, many depend on empirical assumptions about the extent of various Pareto imperfections (including those caused straightforwardly by extant law), the percentage of our society's resources devoted to different kinds of uses (including innovative versus non-innovative QV-investment creation), and, relatedly, the percentage of the resources devoted to marginal uses of given kinds that are withdrawn from various alternative types of uses. At least some of the allocative-efficiency conclusions that follow depend on my guestimates of the relevant parameters. One could argue that, given this fact, I should not include this section at all: that I should restrict myself to explaining why the current consensus on various issues is improperly grounded and why a defensible argument about them would have to be based on a more sophisticated understanding of the factors that affect the allocative efficiency of our current investment and R&D choices as well as on more information about the relevant parameters than is currently available. Nevertheless, I will proceed to rely on various implicit guestimates of the relevant parameters and to comment on the policy-implications of the combination of these assumptions and my theoretical conclusions. If an excuse is required, it would be that, in the real world, the allocative cost of delay often renders it third-best-allocatively-efficient to base decisions on data that is worse than the data that would be TBLE to collect if delay were costless.

The third caveat is a moral caveat. The analyses that follow ignore distribution-related moral issues. As Part I indicated, any complete policy evaluation must consider not only the allocative efficiency of the options reviewed but also their impact on the rights of the parties they affect (whether they are prohibited or required by our rights-commit-ments) and their distributional desirability (when rights-considerations are not decisive) from one or more relevant, legitimate ultimate-value perspectives. Part V will consider only allocative-efficiency issues. It will not address, for example, whether particular patent, copyright, or trade-secret

proposals might violate the discoverer's moral and constitutionally protected property rights (whether they might constitute takings without just compensation in violation of the Fifth Amendment to the United States Constitution, a violation of the Ninth Amendment's general prohibition of violations of rights of the people, or a violation of the Fourteenth Amendment's "privileges and immunities" clause) or whether particular antitrust-law proposals might be undesirable overall even if they are allocatively efficient because they would redistribute income sufficiently undesirably from some relevant, legitimate value-perspective to justify this conclusion.

Having issued these three warnings, Part V will now analyze the implications of Parts II–IV for the allocative efficiency of various antitrust, tax, tort, and patent-copyright-and-trade-secret policy proposals. The analyses that follow will be based on the following four premises for which Parts II-IV argued: (1) from the perspective of allocative efficiency, too many resources are currently devoted to QV investment—regardless of whether the relevant QV investments are non-innovative or technologically or commercially innovative, the marginal QV investments and a significant number of intra-marginal QV investments in each ARDEPPS are allocatively inefficient; (2) from the perspective of allocative efficiency, too few resources are currently devoted to PPR—thus, regardless of whether the unprofitable PPR projects in question that were closest to being profitable would produce discoveries whose use would reduce externalities and whose existence would change the liability of an integrated discoverer/discovery-user but particularly when the relevant projects would have these effects, allocative efficiency would be increased if additional resources were devoted to PPR; (3) from the perspective of allocative efficiency, too few resources are devoted to contemporaneous unit-output production; and (4) although my own lack of expertise in both the law and the relevant business behaviors puts me at a disadvantage in this area, I suspect that in an otherwise-Pareto-perfect world, our current patent-copyright-and-trade-secret (innovation) law would offer certainty-equivalent rewards to particular prospective researchers that are not highly or strongly correlated with the certainty-equivalent allocative value of their prospective research projects (or with the sum of that value and the aggregate distortion in the private cost of the relevant projects).

A. Antitrust Policy Implications

A number of experts have recently argued that economic efficiency would be increased if antitrust exemptions were granted to business arrangements that increased the quantity or proficiency of the R&D their

participants executed.[43] Even if I were persuaded that the vast majority of mergers, joint ventures, and consortia of other types that would allegedly increase the quantity or proficiency of their participants' research-efforts do so[44] or that those arrangements that would actually yield these supposed benefits could be identified at non-prohibitive cost through a litigative process, I would be disinclined to grant the transactions in question antitrust exemptions. Indeed, I would oppose such exemptions even if the only alternative to granting them would be to prohibit all research-promoting or research-efficiency-promoting business arrangements of these kinds that tend to reduce price and non-innovative QV-investment competition on balance. More specifically, I would oppose such exemptions for three reasons that Parts II-IV analyzed: (1) the proposal to exempt from antitrust liability business arrangements that increase the quantity and proficiency of the research in which their participants engage even if the arrangements decrease competition on balance does not distinguish between business arrangements that would increase the quantity of product-research and business arrangements that would increase the quantity of PPR; although increases in the private *proficiency* of both types of research are presumptively allocatively efficient, increases in the *quantity* of product-research will almost certainly tend to decrease allocative efficiency; (2) the proposal ignores the fact that any reduction in price competition that a transaction generates will exacerbate our current tendency to devote too many resources to QV-investment creation (including to product-research) and not enough resources to PPR and unit-output production from the perspective of allocative efficiency; and (3) like virtually all standard economic analyses of the allocative efficiency of pro-price-competition policies, the proposal ignores most of the other types of misallocation that reductions in price competition will generate. More specifically, proponents of these proposals underestimate the magnitude of the allocative inefficiencies the business-arrangements

---

[43] For discussions of proposals that such exemptions be granted, see sources listed in note 2 *supra*.

[44] My skepticism of the relevant exemption proposals is based on my rejection of similar assumptions that underlie the argument that horizontal and conglomerate mergers are virtually always motivated by their capacity to increase their participants' profits by increasing their general organizational allocative efficiency rather than by decreasing the competition they face or enabling them to take advantage of various investor irrationalities or tax-code peculiarities. This claim has been substantially undermined by two later empirical studies. The first study demonstrated that mergers completed in the years between 1950 and 1972 actually decreased the merger partners' market shares relative to those of their non-merging competitors. *See* Dennis C. Mueller, *Mergers and Market Share*, 67 REV. OF ECON. AND STAT. 259, 263–64 (1984). The second study used data on mergers completed between 1950 and 1976 and on sell-offs through 1981 to demonstrate that mergers reduced the rates of return and share-prices of merger partners relative to their counterparts for their non-merging competitors. *See* DAVID RAVENSCRAFT & F.M. SCHERER, MERGERS, SELL-OFFS, AND ECONOMIC EFFICIENCY 217 (1987).

exempted from antitrust regulation that will be caused by reducing price and QV-investment competition.

This last point requires some elucidation. I will restrict myself here to listing the various types of misallocation that reductions in price competition will tend to generate and the various allocative-efficiency effects of reductions in QV-investment competition.[45]

Reductions in price competition will tend to increase many types of misallocation in addition to inadequate-PPR, inadequate-unit-output, and excessive-QV-investment misallocation. First, reductions in price competition will tend to increase two types of consumption-optimum misallocation:[46] (1) unequal-exchange-rate misallocation, which occurs when two consumers of a given pair of goods must incur different relative costs at the margin to buy them and which decreases in price competition increase by increasing the amount of price discrimination, both by increasing the amount of discrimination practiced by best-placed sellers among the buyers they are best-placed to serve and by increasing discriminatory undercutting and retaliation by worse-than-best-placed sellers by increasing the incidence of price-fixing; and (2) by generating a redistribution of income away from the poor, which probably imposes more external costs on those who morally disapprove of such redistributions than external benefits on those who morally (or immorally) approve of them.

Second, in addition to increasing inadequate-PPR misallocation, reductions in price competition cause production-optimum misallocation[47] in four other ways: (1) by increasing price discrimination and hence the amount of misallocation caused by input buyers' facing unequal relative-factor-costs at the margin; (2) by redistributing income away from the poor and thereby increasing the extent to which, from the perspective of allocative efficiency, we underinvest in the children of the poor and in poor adults; (3) by causing undercutting and retaliation that lead to sales of standardized inputs being made by sellers who are not privately-best-placed and who are therefore presumptively not allocatively-best-placed to make them; and (4) by increasing do-it-yourself-labor-*versus*-wage-

---

[45] For a complete treatment of the definition and causes of the various types of resource misallocation, see Richard S. Markovits, *The Causes and Policy Significance of Pareto Resource Misallocation: A Checklist for Micro-Economic Policy Analysis*, 29 STAN. L. REV. 1 (1976).

[46] Consumption-optimum misallocation is present to the extent that a transaction-costless redistribution of the goods actually produced could give its beneficiaries a total equivalent-dollar gain that would exceed the total equivalent-dollar loss it would impose on its victims.

[47] Production-optimum misallocation is present to the extent that some transaction-costless redistribution of factors of production could give its beneficiaries a total equivalent-dollar gain that would exceed the total equivalent-dollar loss it would impose on its victims by making it possible to produce more units of some goods without producing fewer units of other goods.

labor misallocation, which has a production-optimum-misallocation component.

Third, in addition to exacerbating our tendency to devote too many resources to QV-investment creation and not enough to unit-output production, reductions in price competition will tend to increase top-level-optimum misallocation in six different ways: (1) by increasing intra-ARDEPPS relative-unit-output misallocation by increasing price-fixing and the undercutting and retaliation (on non-standardized as well as standardized goods) with which it tends to be associated; (2) by redistributing income away from the poor and thereby increasing the extent to which the poor find it advantageous to misallocate resources both by consuming sub-standard products whose disadvantages they underestimate and by consuming cheaper products whose consumption generates more externalities than would their more expensive alternatives; (3) by increasing labor-leisure misallocation and do-it-yourself-labor-*versus*-wage-labor misallocation, which also has a top-level component, since the relevant worker will not normally be intrinsically indifferent to doing the two types of labor in question; (4) by increasing the amount of misallocation individuals generate by committing crimes (by reducing the attractiveness of their legitimate-behavior options); (5) by increasing X-inefficiency[48] by increasing the incomes of managers and workers, the tax-rate applied to their marginal incomes, and therefore the extent to which the fact that implicit income is not taxed leads them to misallocate resources by substituting for monetary wages untaxed forms of compensation—for example, nicer offices, more attractive secretaries, co-workers against whom they are not prejudiced, less-burdensome work-regimes; and (6) by increasing intra-ARDEPPS QV-investment misallocation[49] by increasing the bias in favor of more-expensive product-variants (imperfections in price competition create such a bias by deflating the extra cost of producing more expensive product-variants).

In short, the advocates of the proposed exemption on which we are now focusing have ignored most if not all of the items in the preceding list.[50] Therefore, I have no doubt that they have substantially underesti-

---

[48] X-efficiency is inefficiency associated with producers' incurring unnecessarily high costs, some of whose components represent a type of relative-unit-output misallocation.

[49] Intra-ARDEPPS QV-investment misallocation is the amount of misallocation that results from an ARDEPPS' creating a less-allocatively-efficient set of products than it could create with the resources it devoted to QV investment.

[50] Unfortunately, the failure of these experts to consider the myriad of ways in which imperfections in price competition misallocate resources is not surprising. Most economists seem to think that pro-price-competition policies increase economic efficiency solely by decreasing relative-unit-output misallocation. In fact, economists do not even do an acceptable job of analyzing the extent to which pro-price-competition policies decrease RUO misallocation. In particular, the existing estimates of these supposed benefits of pro-price-competition policies are deficient in that (1) they try to estimate the effects of anti-trust on relative-unit-output misallocation by estimating the amount of such misallocation currently present in our economy (whereas the relevant figure is the additional amount of

mated the amount of non-research-related misallocation that the exemptions they are recommending would generate by reducing price competition.

I am also confident that the advocates of the exemption with which we are now concerned have underestimated the amount of misallocation that the exemption would generate by reducing non-innovative QV-investment competition. Indeed, since I suspect that these advocates fail to distinguish QV-investment competition from price competition, it would be more accurate to say that they ignored this possibility altogether. Admittedly, my conclusion that too many resources are devoted to QV-investment creation implies that the direct effect of the exemption's tendency to reduce QV-investment competition (the reduction it will yield in non-innovative QV investment) will be allocatively efficient. However, I am confident that any related allocative-efficiency gains will be outweighed by the misallocation that reductions in QV-investment competition cause indirectly (1) by redistributing income away from consumers (who are relatively poorer) and toward shareholders, managers, and workers (who are relatively less poor) and (2) by reducing price competition. Thus, redistributions of income away from (poorer) consumers to shareholders, managers, and workers will tend to misallocate resources (1) by generating direct-income-distribution-value-related externalities, (2) by increasing the extent to which we underinvest in the poor, (3) by increasing the extent of the misallocation the poor generate by committing crimes by more than the reductions decrease the amount of misallocation the better-off generate by committing crimes, (4) by in-

---

such misallocation the economy would generate in the absence of our current pro-price-competition policy or, better yet, the amount by which an allocatively efficient or overall ideal antitrust policy would reduce such misallocation) and (2) the estimates of current relative-unit-output misallocation are too high in that they ignore second best (i.e., the fact that marginal allocative cost exceeds marginal cost) and double-count (in that they count both the misallocation caused by the relative overproduction of some goods and the misallocation caused by the relative underproduction of other goods). A partial bibliography of the relevant literature would include Keith Cowling and Dennis C. Mueller, *The Social Costs of Monopoly*, 88 ECON. J. 724 (1978); Abram Bergson, *On Monopoly Welfare Losses: A Reply*, 65 AMER. ECON. REV. 1024 (1977); R. Carson, *On Monopoly Welfare Losses: A Comment*, 65 AMER. ECON. REV. 853 (1973); Dean A. Worcester, Jr., *New Estimates of the Welfare Loss to Monopoly*, 40 So. ECON. J. 234 (1973); FRANKLIN MICHAEL SCHERER, INDUSTRIAL MARKET STRUCTURE AND MARKET PERFORMANCE (1970); David R. Kamerschen, *An Estimate of the Welfare Loss From Monopoly in the American Economy*, 4 WESTERN ECON. J. 221 (1966); David Schwartzman, *The Burden of Monopoly*, 68 J. POL. ECON. 627 (1960); George J. Stigler, *The Statistics of Monopoly and Merger*, 64 J. POL. ECON. 33 (1956); and Arnold C. Harberger, *Monopoly and Resource Allocation*, 66 PAPERS AND PROC. AM. ECON. ASSOC. 77 (1953). I should add that the double-counting error noticed above may also be made in relation to the type of misallocation on which this Article has primarily focused—that estimates of the amount of misallocation associated with an economy's division of resources among PPR-executing, unit-output-producing, and QV-investment-creating uses should not count both the misallocation associated with the fact that too few resources are devoted to PPR and unit-output production *in toto* and the misallocation associated with the fact that too many resources are devoted to QV-investment creation *in toto*.

creasing the extent of the misallocation the poor cause by consuming substandard goods whose disadvantages they underestimate and externality-generating goods, and (5) by increasing the amount of X-inefficiency managers and workers cause by substituting untaxed forms of wages for monetary wages. Moreover, as the text has indicated, the decreases in price competition that decreases in QV-investment competition will generate will cause many other kinds of resource misallocation as well. On balance, then, I suspect that the decreases in non-innovative QV-investment competition to which the exemption would lead would also misallocate resources.

For all these reasons, I would oppose granting antitrust exemptions to mergers, joint ventures, and consortia that are alleged to increase the magnitude and proficiency of their participants' research-efforts. Moreover, even if I would prefer this exemption to no change in antitrust policy, I would not prefer it to an alternative method of responding to any such efficiencies the relevant business-arrangements would generate, *viz.*, authorizing the Justice Department to sell the right to participate in an on-balance anti-competitive merger, joint venture, or consortium that is alleged to increase the quantity of the participants' PPR or the proficiency of their PPR or product-research for a price that is designed to offset what would otherwise be the distortion in the profitability of the arrangement to them.[51] Since economists have advocated this type of approach in other contexts—most importantly, in the environmental-protection area—I find it somewhat surprising that no one else has recommended this antitrust equivalent to a pollution tax or pollution license.[52]

Of course, it may be that the failure of others to propose such a merger-license scheme reflects their perception that it would be prohibitively expensive to calculate the appropriate fee. It clearly would be complicated to make the relevant calculations. The antitrust enforcement agencies would have to make four separate calculations or sets of calculations in order to determine the distortion-offsetting fee. First, the agency would have to calculate the amount of misallocation that the business-arrangement under review would generate if it had no tendency to increase the quantity or proficiency of its participants' product and production-process research. To do this, the agency would have to predict

---

[51] For previous discussions of this proposal, see Richard S. Markovits, *Monopolistic Competition, Second Best, and* The Antitrust Paradox*: A Review Article*, 77 MICH. L. REV. 567, 606–07 n.71 (1979); *An Ideal Antitrust Law Regime*, 64 TEX. L. REV. 251, 330–32 (1985); *The American Antitrust Laws on the Centennial of the Sherman Act: A Critique of the Statutes Themselves, Their Interpretation, and Their Operationalization*, 38 BUFF. L. REV. 673, 678–773 (1990).

[52] The two sets of policies are analogous because, rather than prohibiting an act whose profitability is distorted, they impose a charge on the relevant choice that is designed to offset the original distortion in its profitability and allow anyone who is willing to pay that charge to engage in the behavior in question.

the effect the arrangement would have on price and QV-investment competition under these circumstances, the effects that these changes in the intensity of competition would have on allocative efficiency, and the amount by which the arrangement in question would increase economic efficiency by increasing its participants' organizational allocative efficiency in ways unrelated to their R&D. Second, the agency would have to calculate the profits the business-arrangement would yield its participants if it would not increase the quantity or proficiency of their R&D. Third, the agency would have to calculate the relationship between (A) the ratio of (i) the allocative-efficiency gains the business-arrangement would generate by increasing the quantity and/or proficiency of its participants' R&D to (ii) the profits the business-arrangement would generate for its participants on these accounts and (B) the magnitude of the private gains just described—a calculation that will be complicated by the fact that the relevant ratio will vary not only with the magnitude of the profits in question but also with the proportions of these gains attributable, respectively, to its increasing the proficiency of their R&D, the quantity of their PPR, and the quantity of their product R&D. And fourth, the agency would have to calculate the fee that the preceding calculations imply would be distortion-offsetting.

A highly simplified numerical example may be useful. Assume that the relevant calculations yielded the following results: (1) absent any effects on the quantity and proficiency of its participants' R&D, a particular merger would misallocate resources by $3,000,000 by reducing price and QV-investment competition; (2) absent any such effects, the merger would increase the participants' profits by $800,000 (*inter alia*, by decreasing price and QV-investment competition); and (3) the ratio of the allocative-efficiency gains the merger would generate by increasing the quantity and/or proficiency of its participants' R&D to the profits it would yield on these accounts would be 3/2, regardless of the size of the relevant profits or the proportion of such profits attributable to each of the three sources listed above. In these circumstances, the fee whose payment would guarantee the allocative efficiency of the business-arrangement in question if both the antitrust authorities and the relevant private actors were sovereign maximizers would be $2,800,000. To see why this is the case, note that (1) to be allocatively efficient, the merger's relevant R&D effects must increase allocative efficiency by infinitesimally more than $3,000,000, (2) the merger would do so only if the profits its R&D effects would yield were at least $2,000,000, and (3) since the merger would increase its participants' profits by $800,000 even if it had no relevant R&D effects, the merger's relevant R&D effects would not generate $2,000,000 in profits and hence $3,000,000 in allocative-efficiency gains unless the merger participants were willing to pay at least $2,800,000 for the right to merge.

Obviously, it would be extremely difficult—indeed, impossible—to execute such a procedure perfectly. The preceding description oversimplified the critical third step, ignored possible merger-partner errors and risk-averseness, and assumed (I think plausibly) that merger-partner license-purchasing decisions would not be affected by strategic incentives related to the merger partners' interactions with the Justice Department and the Congress. Still, I do think that some more-or-less-refined version of this approach would probably be superior either to the proposed exemption or to ignoring the possible tendency of some business-arrangements to increase research-related allocative efficiency (and other types of allocative efficiency) in various ways. Certainly, the practical difficulty of implementing this procedure is not greater than that of its pollution-tax or pollution-license counterparts.

I want to close this analysis of the antitrust-policy implications of Parts II–IV with a more general, positive conclusion. Parts II–IV strengthen the policy case for pro-price-competition policies by demonstrating that such policies will tend to increase allocative efficiency by reducing the amount of resources we devote to QV-investment creation and increasing the amount that we devote to PPR, the production of units of existing goods, and labor.[53]

B. Tax Policy Implications

I will begin once more with the negative implications. As we have seen, Parts II-IV do not favor tax provisions that have been designed to encourage research and investment in general, such as provisions permitting the immediate expensing of research-expenses that do not create physical capital and that have pay-off periods of more than one year,[54] accelerated-depreciation provisions for physical capital,[55] research-cost tax credits,[56] and investment-credit provisions.[57] More specifically, al-

---

[53] Admittedly, Part II's conclusion that we currently devote too many resources to QV-investment creation disfavors pro-QV-investment-competition policies, though as the text indicates, given that decreases in QV-investment competition cause price competition to decrease, pro-QV-investment-competition policies are probably allocatively efficient on balance.

[54] *See, e.g.*, I.R.C. § 174 (providing that research expenses incurred by a business do not have to be charged to a capital account but can be treated as immediate expenses). For a brief discussion of the accidental origin of the provision authorizing research and experimental costs to be expensed, see Calvin Johnson, *Soft Money Investing Under the Income Tax*, 1980 U. ILL. L. REV. 1019, 1078–79 [hereinafter *Soft Money Investing*].

[55] For a discussion of both expensing and accelerated-depreciation provisions, see *Soft Money Investing*, *supra* note 54.

[56] I.R.C. § 41 creates a 20% research tax credit. For a critique of this provision, see Martin Sullivan, *The Research Credit: A Perfect Example of an Imperfect Code*, TAX NOTES 128 (Oct. 11, 1999) (arguing that, though it is conceivable that the private sector's supply of research investment is inefficiently low, the benefits of the tax credit are uncertain because of uncertainty regarding the responsiveness of investment to tax incentives). *See also* Gary Guenther, *The Research and Experimental Tax Credit*, CONGRESSIONAL

though such provisions' tendencies to increase PPR is allocatively efficient, their simultaneous tendencies to increase the amount of resources devoted to both innovative and non-innovative QV investments and to decrease the amount of resources devoted to the production of pre-existing products are undoubtedly misallocative. In fact, I suspect that these tax provisions misallocate resources on balance.

I do think that it would be allocatively efficient (and desirable overall) to use tax policy to change the proportions of our resources devoted to different types of uses. In particular, Parts II–IV suggest that, to increase allocative efficiency in this way, tax provisions will have to be written more selectively to create *de facto* "subsidies" to PPR and unit-output production that render profitable resource-uses of these types that would otherwise be unprofitable and impose *de facto* "penalties" on QV-investment creation that render unprofitable some QV investments that would otherwise be profitable. Indeed, tax law may be able to increase allocative efficiency even more if the relevant provisions distinguish not only among types of resource-uses but also among the ARDEPPSes in which a particular type of research-use is taking place. Thus, even if one did not use tax law to affect the total amount of resources devoted to QV-investment creation, one might use it to increase allocative efficiency (to decrease what I have termed inter-ARDEPPS QV-investment misallocation) by changing the relative amounts of resources devoted to QV-investment creation in different areas of product-space. This conclusion reflects the fact that the Pareto imperfections that the economy currently contains inflate the profitability of the marginal QV investments in different areas of product-space to different extents, which implies that the marginal QV investments in different areas of product-space misallocate resources to different extents. Even if it did not affect the total amount of resources devoted to QV investment, a policy of making the effective rate of taxation on the profits yielded by QV investments in any area of product-space increase with the aggregate positive distortion in the profitability of the marginal QV investment in that area of product-space would increase allocative efficiency by shifting QV investment from ar-

---

RESEARCH SERVICE REPORT No. 96-505 E (June 10, 1998) (arguing that the effectiveness of the research and experimental tax credit is undercut by the fact that the tax credit is not permanent, has a marginal effective rate far below its statutory rate, and is based on an ambiguous definition of qualified research that requires IRS agents to draw difficult distinctions between research that is and is not innovative); William A. Cox, *Research and Experimentation Tax Credits: Who Got How Much? Evaluating Possible Changes*, CONGRESSIONAL RESEARCH SERVICE REPORT No. IB92-039 (June 10, 1996) (arguing that provisions of the tax credit result in varying incentives for different firms and projects—in particular, that firms whose research-spending-to-sales ratios had doubled since the mid-1980s have only half as large a tax incentive to accelerate research and experimentation as do firms with smaller increases, that new, research-intensive firms face only half the incentive motivating less research-intensive new firms, and that capital-intensive projects are less encouraged than are labor-intensive projects).

[57] The United States investment tax-credit was repealed in 1986. *See supra* note 38.

eas of product-space in which the relevant aggregate distortion was higher to areas in which the relevant aggregate distortion was lower.

C. Tort and Environmental Policy Implications

Part III indicated that the most misallocative decisions not to do commercial research are likely to be decisions by integrated potential-researcher/discovery-users who are liable in tort or environmental law only for accident or pollution losses that they are deemed to have caused negligently. This conclusion reflects the fact that, in practice, producers who are potential researchers are never found negligent for failing to try to discover less-accident-and-pollution-cost-prone production-processes or product-variants whose combined production and consumption are less-accident-and-pollution-cost-prone. In fact, in practice, even when the private cost of such a PPR project would clearly be less than the certainty-equivalent private benefits it would be expected to generate by yielding a production-process discovery whose use would reduce accident and pollution losses by more than it increased other types of production costs, the potential researcher's failure to execute the PPR project would not even be assessed for negligence, perhaps because plaintiff-lawyers have implicitly assumed that producers have no duty to do such research. Similarly, even when the private cost of a product-research project would clearly be less than the certainty-equivalent private benefits it would be expected to generate by developing a product whose production and consumption would reduce accident and pollution costs (relative to the accident and pollution costs that would be generated by the production and consumption of the product for which it would be substituted) by more than it would increase other types of production costs and decrease the equivalent-dollar value of the units in question to their consumers, the potential researcher's failure to execute the product-R&D project would not be assessed for negligence.

This deficiency in our current tort law can be corrected in two ways. First and most obviously, by changing negligence law by recognizing a legal duty to execute accident-and-pollution-cost-reducing research-projects whose rejection would fail the normal negligence test. Second, one could correct this deficiency in our current tort law by making producers strictly liable for the accident and pollution losses that the production and (perhaps non-negligent) consumption of their products generate.[58]

This Article also has some more general implications for tort-law reform. In particular, the TBLE approach this Article has described and imperfectly executed implies that the distortions that monopoly and vari-

---

[58] For a more general discussion of the allocative-efficiency case for shifting from negligence to strict liability, see *Negligence Versus Strict Liability*, *supra* note 6.

ous other non-externality Pareto imperfections would create in the profitability of avoidance to various potential injurers would (1) render allocatively inefficient first-best-allocatively-efficient tort or environment law, (2) in some cases, make it allocatively efficient to have negligence determinations turn on comparisons of the relevant allocative costs and benefits of avoidance rather than on the private figures on which the law currently focuses, and (3) favor supplementing tort-damage awards with positive and negative civil fines designed to eliminate the distortion in injurer avoidance-incentives that traditional tort-damage awards would fail to remove in either a negligence or a strict-liability regime.[59]

D. Patent, Copyright, and Trade-Secret Policy Implications

Both the courts and members of Congress claim that the current law does not give sufficient protection to creators of knowledge or even of images and entertainment forms. Although these conclusions could reflect property-right-related moral-rights notions or might be little more than political rhetoric, these days such claims might well reflect a sincere belief that the incentives that the law provides potential "creators" are sub-optimal from the perspective of economic efficiency. Such a belief appears to underlie a variety of case-outcomes—e.g., holdings that the copyright laws protect architectural plans and drawings, including design elements based on those plans and drawings; that the postcard of a city skyline infringes the trademark rights of a plaintiff's recognizable building; that a celebrity's right to publicity includes not only the use of her name, but also the use of her picture, voice, and overall image as well as the right to prevent someone from looking or sounding too much like her; that the copyright law of fair use is far more limited than it formerly was, etc. Members of Congress also make reference to this argument when considering legislation aimed at "reversing" judicial decisions against "creators" by protecting databases or boat-hull designs as well as when advocating legislation that would extend an author's copyright protection from 50 to 70 years (though, admittedly, the present-value effects of such legislation would be small).[60]

---

[59] For a detailed discussion of these issues, see Richard S. Markovits, *Monopoly and the Allocative Inefficiency of First-Best-Allocatively-Efficient Tort Law: The Whys and Some Therefores*, 46 CASE W. RES. L. REV. 313, 427–30 (1996).

[60] For the citations to and a fuller discussion of the cases summarized in the text as well as an analysis of the general trend they manifest, see generally Mark Lemley, *Romantic Authorship and The Rhetoric of Property*, 75 TEX. L. REV. 873, 898–902 (1997) (book review). For an unusually sophisticated analysis of the allocative-efficiency case for granting, lengthening, or broadening copyright protection, see Glynn S. Lunney, Jr., *Reexamining Copyright's Incentives-Access Paradigm*, 49 VAND. L. REV. 483 (1996). For recent Congressional legislation dealing with some of these issues, see Digital Millennium Copyright Act, 17 U.S.C. §§ 302, 512, and 1201–05 (1998).

To the extent that these judicial decisions and proposed bills reflect a belief that current "innovation law" provides creators with allocatively-suboptimal incentives, they run counter to Part II's conclusion that, from the perspective of allocative efficiency, too many resources are currently devoted to creating QV investments, including QV investments that are innovative.

Of course, a complete allocative-efficiency analysis of proposals to extend patent or copyright protection to additional types of discoveries or to lengthen or broaden such protection would have to consider not only their impact on the allocative efficiency of R&D decisions but also their impact on the amount of misallocation that we generate by underusing and inventing around protected discoveries from the perspective of allocative efficiency, their impact on the allocative transaction costs discoverers and discovery-users generate when contracting with each other and engaging in discovery-use-related legal disputes, as well as their impact on the net externalities that IP law generates by pleasing or displeasing those whose distributional values make them concerned about its distributional effects.

To be honest, although I admit to great ignorance in this area, I would be surprised if our current system did not cause substantial misallocation by leading potential researchers to make allocatively inefficient R&D choices (by distorting the certainty-equivalent profitability of most product and production-process research-decisions), by causing discoveries to be underutilized and invented around, and by generating huge allocative transaction costs. Worries about government inefficiency and corruption make me hesitate to propose replacing our current system with a combination of direct government payments to successful discoverers and a requirement that discoverers reveal and allow free use of their discoveries by domestic users. However, my suspicion that our present system is extraordinarily economically inefficient in comparison with the ideal makes me willing to consider the possible advantages of this alternative (whose practicability could be increased if the government payments were made over time since the passage of time would provide further information on the *ex post* value of having a discovery made earlier, which would improve our estimate of the relevant research-effort's *ex ante* certainty-equivalent allocative value).

Two final legislative proposals respond to Part III's conclusion that, from the perspective of economic efficiency, we currently devote too few resources to production-process research. The first proposal relates to one factor not previously discussed that exacerbates this tendency to do too little PPR: the difficulty that production-process patent-holders have in determining whether producers of products to whose production their discovery relates are using their process. Since production takes place behind closed doors, it is far more difficult to discover a production-process-patent violation than a product-patent violation. Congress has

already responded to this difficulty when the possible violator is a foreign producer who may have taken advantage of his host country's more lax patent protection. When there is a substantial likelihood that a foreign producer has violated the American patent laws by using a patented production process and, despite reasonable efforts, the patent-holding plaintiff has not been able to determine whether a violation has taken place, a court is authorized to prohibit the importation of the good in question.[61] Admittedly, since production-process patent-holders will have less difficulty gaining access to a potential domestic violator's production facilities, such a *de facto* presumption of illegality may not be warranted when the defendant is a domestic producer. However, it may be economically efficient to increase the profitability of PPR by enabling production-process patent-holders to enforce their patent rights in other ways.

The second proposal is designed to increase the profitability of doing production-process research by facilitating the patent-application process. Congress has passed legislation that encourages both production-process R&D and product R&D by giving production-process and product discoverers a grace period that allows them to apply for a patent up to twelve months after they have disclosed their discovery to the public.[62] By enabling discoverers of possibly patentable products or production processes to delay their patent-application decisions until they have learned more about the commercial value of their discovery by marketing it for a year, the grace period increases the profitability of product and production-process research by increasing the private cost-effectiveness of the discoverers' patent-application choices. In particular, the provision enables such discoverers to avoid making patent applications when it would prove to be *ex post* unprofitable for them to do so and enables them to make profitable adjustments in the patent-application efforts they do make, for example, in the breadth of the patents they try to secure.

Unfortunately, there is some reason to believe that the current "grace period" provision may not be more economically efficient than a policy giving no grace period to innovators of any kind or a policy of giving production-process discoverers but not product discoverers a grace period. Moreover, there is considerable reason to believe that economic efficiency would be increased if production-process discoverers did not have to disclose their innovations to qualify for a grace period. From the perspective of economic efficiency, the current provision is problematic because it probably increases the profitability of product research by more than it increases the profitability of production-process research. This conclusion does not reflect any assumption that production-process discoverers are more able than product discoverers to predict the com-

---

[61] *See* 35 U.S.C. § 295 (1998).

[62] *See* 35 U.S.C. § 102(b) (1998). *See also* DONALD S. CHISUM, CHISUM ON PATENTS § 602(1) (2000).

mercial value of their respective types of discoveries before marketing them. Instead, it reflects two other premises. The first relates to the positive law: although case law suggests that the act of marketing new products to the public will always be deemed to satisfy the "public disclosure" requirement,[63] it also suggests that production-process innovators who use their discovered production-process themselves will not be deemed to have made the required public disclosure by marketing the product they use their discovery to produce. The second premise relates to the possibility that a production-process discoverer who would otherwise find it most profitable to use his discovery exclusively himself rather than to sell it or license someone else to use it might satisfy the "public disclosure" requirement by revealing his discovery directly to the public. For two reasons, such an explicit disclosure might cost a production-process discoverer more by facilitating potential rivals' inventing around any patent he should obtain than a product innovator's marketing his new product would cost the product innovator in this way: (1) the direct revelation of the innovative production-process would probably reveal more about the discovery in question than would the marketing of the new product (given that reverse engineering is always expensive and may be impracticable or impossible, the direct revelation of the production-process discovery is more likely to make copycat activity profitable than would the sale of a new product, *ceteris paribus*), and (2) economies of scale relative to the extent of demand may be more likely to preclude potential rivals from profiting by inventing around product innovations than by inventing around production-process innovations.

In any event, to the extent that the above two conditions are fulfilled, the "grace period" provision—35 U.S.C. § 102(b)—may cause economic inefficiency by shifting resources that are especially well-suited to technological innovation (most importantly, scientists and engineers) from production-process research to product research. In fact, § 102(b) may actually be misallocative on balance. Admittedly, a complete or even a third-best-allocatively-efficient analysis of § 102(b) would have to take several of its other impacts into account. Although it is beyond the scope of this Article to execute even a third-best analysis here, the following three additional considerations are probably most important in this connection. First, if my conclusion that we currently have too much product innovation from the perspective of economic efficiency holds even for product research that yields patentable products, § 102(b) will tend to generate economic inefficiency by causing additional resources to be moved from unit-output production, leisure production, and non-

[63] *See, e.g.*, Metallizing Engineering Co. v. Kenyon Bearing & Auto Parts Co., 153 F.2d 516 (2d Cir. 1946) (L. Hand, J.) (holding that the sale of a product made by a secret process constituted public use of the process for purposes of the Section 102(b) bar on obtaining patent on inventions already in public use for more than one year); *see also* CHISUM, *supra* note 62 at § 6.02(5)(b).

patentable QV-investment creation to R&D that may yield a patentable product. Second, to the extent that production-process researchers would otherwise find it profitable to use their own discoveries exclusively themselves as opposed to selling them or licensing their use and to the extent that such researchers would find it costly (for reasons related to potential rivals' inventing around their patents) to reveal their production-process discoveries directly to the public, § 102(b) will tend to increase economic efficiency less than one would otherwise suppose by shifting resources to production-process research that may yield patents. Third, and in the other direction, I suspect that § 102(b) will tend to increase economic efficiency by inducing production-process discoverers who would otherwise have used their innovation exclusively themselves to sell it to others or to license others to use it.

As already suggested, I do not know whether, as currently written and applied, § 102(b) increases or decreases economic efficiency on balance. However, I am convinced that, if a grace period is to be made available to both product innovators and production-process innovators, its availability to production-process discoverers should not depend on their disclosing their discovery to the public either explicitly or by selling their discovery or the right to use it. If the law's goal is to increase economic efficiency, it should entitle production-process innovators to trigger a grace period by using their production-process discovery themselves. More specifically, such a revision of § 102(b) would increase economic efficiency (1) by removing the feature of the current provision that misallocates resources from production-process research to product research and (2) by increasing the extent to which resources are allocated to production-process research from resource-uses other than technologically innovative product R&D by increasing the profitability of production-process R&D, though admittedly the revision in question might reduce economic efficiency (3) by eliminating the feature of the current provision that tends to induce production-process innovators to sell their discovery to others or license others to use their discovery.

One further possibility should be raised: the most economically efficient "grace period" provision might entitle production-process discoverers but not discoverers of potentially patentable products to a grace period. Admittedly, this revision would tend to decrease economic efficiency *inter alia* by increasing the allocative transaction costs that product discoverers generate when applying for patents both directly and indirectly by inducing their potential rivals to oppose their efforts and requiring the government to process and finance the processing of their claims. However, if I am correct about the current aggregate distortions in the profitability of marginal product-research and PPR projects that might lead to the issuance of patents, a decision to entitle production-process discoverers but not product discoverers to a grace period would tend to increase economic efficiency both by shifting scientists and engi-

neers from product research to production-process research and by shifting resources from product research to various other kinds of economic uses. Once more, I am not sure whether such a decision would increase economic efficiency on balance. However, if it would, the fact that it would be economically inefficient in a world that had no other Pareto imperfections would be irrelevant to its actual economic efficiency if, for political or other reasons, these other imperfections could not be altered sufficiently to critically affect the aggregate distortions that this discussion assumes makes the option under consideration economically efficient.

E. Patent, Copyright, Trademark, and Trade-Secret Law Implications

Recent scholarship suggests that judicial determinations of the breadth of patent and copyright protection have been substantially influenced by the judges' conclusions about the allocative efficiency of broadening the protection given.[64] Traditionally, legal scholars who are conversant with economics[65] and economists[66] have assumed that broader protection will tend to increase allocative efficiency by increasing the "incentive to invent" but will tend to decrease allocative efficiency by reducing to sub-optimal levels the extent to which discoveries are used, by increasing the allocative transaction costs generated by the use of discoveries by making it necessary for non-discoverers to purchase the right to use them, and by encouraging non-discoverers to expend resources to invent around patent protection when, "incentive to invent" issues aside, it would have been more allocatively efficient for them to make free use of the discovery. This Article has pointed out that this traditional analysis is misleading insofar as it assumes that, *ceteris paribus*, the allocative efficiency of our "invention" efforts will be increased by increases in the profitability of R&D or artistic creation. In particular, this Article has raised the possibility that, from the perspective of allocative efficiency, we may already have too much R&D and artistic creation of certain types. Obviously, if this is the case, broadening patent or copyright protection will increase misallocation by increasing the allocative excessiveness of the investments we make in the relevant types of research or artistic creation as well as by restricting the use of the discoveries that are made and increasing various types of allocative transaction costs.

---

[64] *See* Dam, *supra* note 1, at 258–61 (arguing that courts have addressed the breadth of patent protection through the application of various patent-law rules and principles including the doctrine of equivalents; rules addressing price-fixing, tie-ins, and other patent misuses; rules requiring novelty, utility and nonobviousness that sometimes serve to invalidate specific patent claims; and rules permitting the patenting of new uses of, and improvements to, existing patented products).

[65] *See id.*

[66] *See* Machlup, *supra* note 1.

More generally, Parts II and III of this Article lay out the kind of analysis that lawyers in patent-or-copyright-breadth cases should use to persuade judges of the allocative efficiency of giving a discoverer or creator broader or narrower patent or copyright protection. At the most crude level, these Parts imply that arguments for broad protection will tend to be stronger when the discovery is a production-process discovery than when it is a product discovery, at least if a proper interpretation of the law makes relevant the allocative efficiency of the protection awarded. At a more refined level, these Parts delineate and explain the relevance of all the facts that affect the allocative efficiency of broadening or narrowing the patent/copyright protection a discoverer/creator is awarded.

The analysis of Part II also suggests that it may be economically efficient to make the requirements for eligibility for trademark protection more stringent. For example, Part II may favor recent Supreme Court decisions holding that a design or configuration can be given trademark protection only if (1) consumers associate the relevant design or configuration with its original source[67] and (2) the design or configuration is not functional.[68]

CONCLUSION

This Article was designed to achieve two goals. Its primary and more fundamental goal was to explain and exemplify the type of allocative-efficiency analysis that I think must be employed in policy-evaluation contexts—the type of analysis I have called third-best-allocative-efficiency analysis. A complete third-best-allocative-efficiency analysis of a single policy or set of alternative policy-options would delineate all the various possible types of resource misallocation the policy or policies in question would affect; identify all the individual Pareto imperfections that would cause each such type of misallocation on its own; develop formulae that would relate the amount of misallocation of each type that is present in the economy to the various Pareto imperfections that can cause each; combine this theoretical analysis with empirical estimates or guestimates of the various relevant Pareto imperfections pre-policy and post-policy to estimate the allocative efficiency of collecting additional data on the relevant parameters; collect the additional data whose collection seems to be allocatively efficient, given its cost, accuracy, and relevance to the allocative efficiency of the policies under consideration and various other policies until the continuously revised estimates of the allocative efficiency of further data-collection indicate that data-collection should stop; and then examine the implications of the

---

[67] *See* Wal-Mart v. Samara Brothers, 529 U.S. 205 (2000).
[68] *See* Traffix Devices, Inc. v. Mktg. Displays, Inc., 532 U.S. 23 (2001).

executed theoretical and empirical work for the allocative efficiency of the policy or policies in question.

Obviously, this Article's third-best-allocative-efficiency analyses of the antitrust, tax, tort-and-environmental, and patent-copyright-and-trade-secret policies it considered were very partial and preliminary. However, I hope that they were sufficiently complete to reveal the superiority of third-best-allocative-efficiency analysis to (1) the unsubstantiated pronouncements that play major roles both in contemporary political discussions of why and how we should encourage additional investment in general and R&D in particular and in microeconomic-policy debates in general, (2) the first-best-allocative-efficiency analyses that dominate the law-and-economics field, (3) prohibitively expensive, complete second-best-allocative-efficiency analyses (which ignore the inevitable cost and inaccuracy of data and analysis and which, not surprisingly, no-one has ever come close to executing), and (4) the kinds of partial second-best-allocative-efficiency analyses from which much can be learned but which ignore too many types of resource misallocation, too many types of Pareto imperfections, and too many types of distorting feedback-relationships to be allocatively efficient in our highly interconnected and highly Pareto-imperfect world.[69]

The second goal of this Article was to use the results of my admittedly incomplete third-best-allocative-efficiency analysis (1) to criticize the assumption that, from the perspective of allocative efficiency, Western economies now devote too few resources to product and production-process research and, indeed, to investment of all kinds as well as the current policy consensus in favor of antitrust, tax, and patent-copyright-and-trade-secret laws that will encourage additional investment of all kinds, especially in product and production-process research, development, and commercialization and (2) to develop alternative policy-proposals that reflect my own preliminary conclusion that, from the perspective of allocative efficiency, we currently devote too many resources to creating quality-or-variety-increasing investments and not enough resources to executing production-process research and producing units of products already in production.

---

[69] For a more detailed discussion of The General Theory of Second Best, the varieties of third-best analyses that can be executed, and the reasons why economists and economist-lawyers persist in ignoring Second-Best Theory, see Richard S. Markovits, *Second-Best Theory and Law & Economics: An Introduction*, 73 CHI.-KENT L. REV. 3 (1998). For a critique of some standard arguments with which economists attempt to justify their failure to take appropriate account of Second-Best Theory, see Richard S. Markovits, *Second-Best Theory and the Obligations of Academics: A Reply to Professor Donohue*, 73 CHI.-KENT L. REV. 267 (1998).

ARTICLE

ERRATIC BY DESIGN: A TASK ANALYSIS OF PUNITIVE DAMAGES ASSESSMENT

DAVID A. SCHKADE*

Critics of punitive damage awards have characterized them as unpredictable, erratic, or arbitrary. This Article argues that the design of the punitive damages decision may make the current system prone to erratic awards. Comparing current practice in the punitive damages system to simple principles of task design reveals specific strengths and weaknesses. The most serious weakness is that one of the tasks jurors are required to perform—assessing a specific dollar amount of punitive damages—is one that they cannot possibly be expected to perform well. It is jurors' valiant but doomed attempts to perform this poorly designed task that produce erratic awards that fail to treat similarly situated parties similarly. The final Part discusses possible reforms based on this analysis.

I. INTRODUCTION

Juries often surprise observers with the punitive damage awards they assess in carefully argued civil trials. Some awards are surprisingly high, some are surprisingly low.[1] Many observers find these outcomes quite dubious, and punitive damage awards have been characterized as unpredictable,[2] erratic,[3] random and capricious,[4] or worse. Some have questioned the suitability of jurors as decision-makers for the assessment of punitive damages, and they have expressed concerns that jurors may be biased or irrational.[5] Why do these anomalies plague the civil justice

---

* Herbert D. Kelleher Regents Professor, McCombs School of Business, University of Texas, Austin. Ph.D., Carnegie Mellon University, 1985; M.S., Carnegie Mellon University, 1982; M.B.A., University of Texas, Austin, 1980; B.A., University of Texas, Austin, 1978. Daniel Kahneman, Cass Sunstein, and Alison Davis-Blake provided helpful comments on an earlier draft. The writing of this Article was supported in part by Exxon Corporation and also by an appointment as a Visiting Senior Research Scholar in the Woodrow Wilson School of Public and International Affairs at Princeton University.

[1] For an example of both, see BMW of N. Am., Inc. v. Gore, 517 U.S. 559, 559 (1996).

[2] *See* Jonathan Karpoff & John Lott, *On the Determinants and Importance of Punitive Damage Awards*, 42 J.L. & ECON. 527, 571 (1999). *See also, e.g.*, A. Mitchell Polinsky, *Are Punitive Damages Really Insignificant, Predictable, and Rational?*, 26 J. LEGAL STUD. 663, 663–64 (1997); Peter Huber, *No-Fault Punishment*, 40 ALA. L. REV. 1037, 1049 (1989).

[3] Daniel Kahneman, David A. Schkade & Cass R. Sunstein, *Shared Outrage and Erratic Awards: The Psychology of Punitive Damages*, 16 J. RISK & UNCERTAINTY 49, 49 (1998).

[4] CONTRACT WITH AMERICA: THE BOLD PLAN BY REP. NEWT GINGRICH, REP. DICK ARMEY, AND THE HOUSE REPUBLICANS TO CHANGE THE NATION 154 (Ed Gillespie & Bob Schellhas eds., 1994).

[5] *See, e.g.*, Reid Hastie & W. Kip Viscusi, *What Juries Can't Do Well: The Jury's Per-*

system?[6] Is it because jurors are foolish or disingenuous or inherently flawed?

It should come as no surprise that such anomalies occur in the difficult task of determining an appropriate amount of punitive damages. This Article argues that the design of the civil legal system makes it prone to erratic awards. Jurors' punitive damage awards surprise us because the system places superhuman expectations upon them. They are given unlimited discretion but only limited guidance in deciding an amount of punitive damages. They are kept in the dark about factors critical to the determination of an amount, yet are expected to be as wise (or wiser) than those in the light. As a result, assessing an amount of punitive damages is a poorly designed task that may be impossible in principle for jurors to perform consistently or accurately.[7]

This Article takes a task design perspective on punitive damages. Viewing the current regime as a system for making decisions provides an organizing principle for analyzing and understanding why the task of assessing an amount of punitive damages is so difficult for jurors. The erratic character of these awards can then be seen as a predictable consequence of the design characteristics of the punitive damage award decision. By comparing practice in the current punitive damages regime to five simple design principles, specific strengths and weaknesses are revealed. An understanding of these strengths and weaknesses provides a basis for the choice of reforms.

The remainder of the Article is organized as follows. Part II analyzes the task of assessing punitive damages using five simple design principles from management theory. Part III relies on these insights to identify promising (and not so promising) avenues for reform.

II. Principle and Practice in Punitive Damage Assessment

The punitive damages system serves several purposes, principally to punish and deter non-criminal actions that society deems objectionable.[8]

---

formance as a Risk Manager, 40 Ariz. L. Rev. 901, 909–13 (1998).

[6] Of course there are many defenders of the current system. *See, e.g.*, Theodore Eisenberg et al., *The Predictability of Punitive Damages*, 26 J. Legal Stud. 623 (1997) (claiming correlation between compensatory and punitive damages); Marc Galanter & David Luban, *Poetic Justice: Punitive Damages and Legal Pluralism*, 42 Am. U. L. Rev. 1393 (1993) (arguing punitive damages are a worthy form of "civil-like" punishment).

[7] For our purposes here, we shall assume that there are indeed many surprising awards. On the unpredictability of punitive awards, see Cass R. Sunstein, Daniel Kahneman & David Schkade, *Assessing Punitive Damages,(with Notes on Cognition and Valuation in Law)*, 107 Yale L.J. 2071, 2075–77 (1998); George L. Priest, *Punitive Damages Reform: The Case of Alabama*, 56 La. L. Rev. 825, 826–30 (1996); Polinsky *supra* note 2, at 672–74; Karpoff & Lott, *supra* note 2, at 571. The focus will be on why this is so, and whether faulty jurors are to blame.

[8] Jeff D. Brecht, *Oregon's Procedural Due Process and the Necessity of Judicial Review of Punitive Damage Awards*: Honda v. Oberg*: Stop the Insanity!*, 15 N. Ill. U. L. Rev.

Here we will view it as a system for making decisions, focusing in particular on the task of assessing a specific dollar amount of punitive damages. In principle, how could the task of assessing punitive damages be designed to make good decisions? How does practice in the current system measure up to these principles? This Part presents five simple principles for the design of tasks and organizations, and uses them to analyze current practice in assessing an amount of punitive damages:

Hire the best people;
Give them good training;
Give them the right information;
Give them incentives to perform well;
Provide a review process for exceptional cases.[9]

The Article first explains the logic underlying each principle, and then compares it with current practice in the punitive damage regime, with particular emphasis on the task of determining a specific dollar amount. Table 1 summarizes the principles and associated current practice.

Principle: Hire the best people. Perhaps the most basic principle is to get the best decision-makers you can find, those people who have extensive knowledge, training, and experience in making a certain type of decision.

Hiring good people and getting out of their way is a sound and commonly accepted maxim among managers and administrators.[10] It is self-evident that the more capable and more experienced the people in a system are, the better the system will function. One important consequence of greater skill and experience is that many potentially problematic situations or events either never occur in the first place, or are handled quickly and never become exceptions that require intervention by a higher authority.[11] This avoids the problem of placing excessive decision-making loads on the hierarchy of the organization, which in the legal system translates into an overloaded appeals process and clogged court dockets.

Practice: Take whoever you can get, regardless of qualifications. Under the current punitive damages regime, decisions are made by low-paid, randomly selected draftees, who after the filtering of the *voir dire* process, most likely have little relevant knowledge, training or experience, in either the law or the subject matter of the case.

377, 382 (1995).

[9] *See* Jeffrey Pfeffer, THE HUMAN EQUATION: BUILDING PROFITS BY PUTTING PEOPLE FIRST 64–98 (1998) (summarizing and expanding on these and related design principles). These principles are to some extent widely shared and self-evident beliefs among management practitioners and scholars. *See, e.g., id.* at 64.

[10] *Id.* at 69–80.

[11] *See id.* at 85.

How are punitive damage jurors selected? Most English-speaking adults in the United States are eligible to be selected for jury service, although the specific criteria vary from state to state.[12] Most states choose randomly from voter registration or driver's license lists to select individuals for jury service.[13] However, only a tiny minority of jury-eligible citizens ever serve on a jury in an actual trial.[14] Since most of those who are selected serve on criminal juries or on civil juries in which punitive damages are never considered, only a small minority of this tiny group ever serves on a jury that determines a punitive damage award.[15] Thus, it is a rare coincidence indeed when a punitive damage jury contains even one person who has any experience with deciding an amount of punitive damages.

In addition, trial lawyers are generally averse to empanelling jurors with too much knowledge or expertise relevant to the case (including experience with making punitive damage decisions), presumably because this leaves them less open-minded about the evidence and arguments to be presented.[16] Consequently, in the *voir dire* process, most such potential jurors are rejected by one counsel or the other, or even by the judge.[17]

More broadly, for over two centuries there has been a trend in the law away from empanelling jurors who have any relevant knowledge about a case.[18] In part, this was a response to concerns about the objectivity and fairness of people who are too close to the case, or who have other agendas. In the early days of the modern jury, knowledge of the case was actually a criterion used to select jurors.[19] Instead, we now rely on evidence presented by advocates to a "blank slate" of jurors so that

---

[12] BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, STATE COURT ORGANIZATION 1998, at 261 (2000); JAMES J. GOBERT & WALTER E. JORDAN, JURY SELECTION: THE ART AND SCIENCE OF SELECTING A JURY, § 6 (2d ed. 1990).

[13] *See* BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, *supra* note 12, Table 39, at 263–68 for specific requirements by state.

[14] It has been estimated from a study of the Chicago courts over the period 1959 to 1979 that the probability in any year of a Chicago citizen serving on a jury that is actually called to render a verdict (any verdict) was .0038, equal to service once every 260.2 years. George L. Priest, *The Role of the Civil Jury in a System of Private Adjudication*, 1990 U. CHI. LEGAL F. 161, 188–89.

[15] William Glaberson, *Juries, Their Powers Under Siege, Find Their Role Is Being Eroded*, N.Y. TIMES, Mar. 2, 2001, at A1 (explaining that punitive damages are awarded in only about 4% of civil cases that go to trial).

[16] *See* Graham C. Lilly, *The Decline of the American Jury*, 72 COLO. L. REV. 53, 61–66 (2001); *see also* Franklin Strier, *The Educated Jury: A Proposal for Complex Litigation*, 47 DEPAUL L. REV. 49, 72–74 (1997).

[17] *See* Lilly, *supra* note 16, at 61–66; *see also* Strier, *supra* note 16 at 72–74.

[18] *See* Stephen Landsman, *The Civil Jury in America: Scenes from an Unappreciated History*, 44 HASTINGS L.J. 579, 605–10 (1993).

[19] Strier, *supra* note 16, at 58 ("At the inception of the modern jury, all juries were special in that they were purposely chosen for their special knowledge of the facts in dispute.") However, the practice of using people with knowledge of the case as jurors was apparently in decline as early as the seventeenth century. See Landsman, *supra* note 18, at 586.

they will rely only on the information presented to them, thus minimizing the impact of pre-existing beliefs.

What then are the strengths of this selection procedure? To the extent that the civil justice system has populist objectives, this method does have its virtues. It is clear that the availability of the court system to all citizens, coupled with the symbolic significance of being judged by one's peers is an important asset worthy of protection. In addition, since community sentiment plays a key role in punitive damages, ordinary citizens are logical representatives.[20] Further, jurors appear to be capable of reliably assessing the relative outrageousness of different transgressions within the same category of cases.[21] What is less clear, however, is whether expressing this relative outrage by estimating the specific numerical amount of punitive damages required to serve the twin goals of (a) punishing the defendant and (b) deterring such conduct is a suitable role for a random selection of people who are virtually certain to be amateurs at this difficult task.[22]

Principle: Give them good training. If the system cannot get people with skill and experience, perhaps it can give jurors specific instructions and clear criteria, so that if they follow this guidance they can still perform well.

That good training can enhance performance is such a commonplace maxim results in people taking this principle for granted.[23] Systematizing, documenting, and communicating the knowledge necessary to perform a task should make it easier to perform well. The availability of excellent training for a task or occupation probably reduces to some extent the significance of whether or not the most talented and experienced people can be hired.

Practice: Training is brief and instructions are vague. In practice, the training given to punitive damage jurors leaves much to be desired. It comes primarily in the form of instructions provided by the judge. Although there are some differences in basic punitive damages instructions across states, they are generally quite similar.[24] Typical punitive damages

---

[20] *See* Landsman, *supra* note 18, at 592–605 (on the civil jury as the primary means of citizen participation in governance); *see also* Nancy S. Marder, *Deliberations and Disclosures: A Study of Post-Verdict Interviews of Jurors,* 82 IOWA L. REV. 465, 472–73 (1997) (discussing the symbolic value of the jury as the voice of the community).

[21] *See* David Schkade, Cass Sunstein and Daniel Kahneman, *Deliberating About Dollars: The Severity Shift*, 100 COLUMBIA L. REV. 1139, 1146–47 (2000).

[22] Brecht, *supra* note 8, at 382–84.

[23] PFEFFER, *supra* note 9, at 85–90.

[24] Here is a typical instruction:

> If you find from the evidence that [the defendant] is guilty of wanton, willful, malicious or reckless conduct that shows an indifference to the rights of others, then you may make an award of punitive damages in this case.
>
> In order for the conduct of the defendant to constitute willfulness or wantonness, his/her acts must be done under circumstances which show that he/she was aware

instructions and standards given to jurors consist mostly of broad and somewhat vague principles rather than a set of concrete actions to follow or specific criteria to apply. Specifically, there is almost no guidance whatsoever about the appropriate amount of punitive damages.[25]

The Supreme Court has stated that the instructions given by judges can be grossly inadequate.[26] In *BMW v. Gore*, Justice Breyer spoke in some detail about the failure of the law to limit jury discretion with clear criteria about the grounds of awards:

> The standards the Alabama courts applied here are vague and open ended to the point where they risk arbitrary results. In my view, although the vagueness of those standards does not, by itself, violate due process . . . it does invite the kind of scrutiny the Court has given the particular verdict before us. . . . This is because the standards, as the Alabama Supreme Court authoritatively interpreted them here, provided no significant constraints or protection against arbitrary results.[27]

Justice O'Connor, in her dissenting opinion in *Pacific Mutual Life Insurance v. Haslip*, argues that the inadequacy of instructions on punitive damages gives juries "unchanneled, standardless discretion,"[28] and goes on to say:

> States routinely authorize civil juries to impose punitive damages without providing them any meaningful instructions on how to do so. Rarely is a jury told anything more specific than "do what you think best." . . . In my view, such instructions are so fraught with uncertainty that they defy rational implementation. Instead, they encourage inconsistent and unpredictable re-

---

from his/her knowledge of existing conditions that it is probable that injury would result from his/her acts and omissions, and nevertheless proceeded with reckless indifference as to the consequences and without care for the rights of others . . . .

It is not necessary to find that the defendant deliberately intended to injure the plaintiff. It is sufficient if the plaintiff proves by the greater weight of the evidence that the defendant intentionally acted in such a way that the natural and probable consequence of his/her act was injury to the plaintiff.

RONALD W. EADES, INSTRUCTIONS ON DAMAGES IN TORT ACTIONS §§ 2-6, 2-7, 2-8 (4th ed. 1998).

[25] *See generally id.* at 99–161 (explaining example of punitive damage instructions from a California court).

[26] *See BMW*, 517 U.S. at 588–92 (1996) (Breyer, J. concurring).

[27] *Id.*

[28] Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 46 (1991) (O'Connor, J., dissenting).

sults by inviting juries to rely on private beliefs and personal predilections.[29]

Jurors often respond to this ambiguity by failing to follow their instructions correctly, or by simply ignoring them altogether, instead substituting their own intuitive criteria.[30] Indeed, there is evidence that jurors serving on cases involving punitive damage awards may recall little of their instructions even immediately after deliberations.[31] Nevertheless, jurors' non-dollar moral intuitions about the relative outrageousness of a defendant's actions appear to be widely shared and reasonably reliable, despite this ambiguity in instructions.[32]

Significantly, punitive damage instructions usually contain no references to even a range of dollar amounts, despite the fact that assessing a specific dollar amount is the task.[33] With unlimited discretion but little guidance in choosing an amount, jurors are left to grasp at straws in their struggle to identify some relevant standard or point of reference for translating their outrage into dollars. As a consequence, different juries appear to employ widely differing standards for assessing a specific amount of punitive damages. For example, interviews with the California jurors who assessed a $4.9 billion award against General Motors suggest that the number came from an estimate of the defendant's advertising budget for the year.[34] In another case, which was appealed to the United States Supreme Court, the jury appears calculated a dollar amount as a function of the total worldwide unit sales of the defendant.[35] It is not clear why the appropriate punishment would be measured by the company's advertising budget on the one hand and by its worldwide sales on the other.

Another consequence of this lack of guidance is a susceptibility to suggestion. It has been well documented that when making numerical estimates of goods or objects that are difficult to assess, people are strongly influenced by whatever numbers happen to be easily available, often regardless of their relevance.[36] These numbers are referred to as

---

[29] *Id.* at 42–43.

[30] Reid Hastie, David A. Schkade & John W. Payne, *A Study of Juror and Jury Judgments in Civil Cases: Deciding Liability for Punitive Damages*, 22 LAW & HUM. BEHAV. 299–304 (1998).

[31] See Strier, *supra* note 16, at 51–52; *see also* Hastie, Schkade & Payne, *supra* note 30, at 294–95.

[32] *See* Schkade, Sunstein & Kahneman, *supra* note 21, at 1146–47.

[33] *See* Haslip, 499 U.S. 1, 46 (O'Connor, J., dissenting); *see also* EADES, *supra* note 24, at 106, 110.

[34] *Jury Orders GM Corp. to Pay $4.9 Billion to Victims of Crash*, BALT. SUN, July 10, 1999, at 3A.

[35] Ford Motor Co. v. Ammerman, 705 N.E.2d 539, 559 (Ind. Ct. App. 1999).

[36] Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1128–30 (1974).

"anchors."[37] In the context of punitive damages, one of the few dollar figures available is that requested by the plaintiff: the amount of damages suggested by the plaintiff's counsel in closing arguments.[38] The plaintiff's request has been shown to have a substantial influence on the amount awarded—the more you ask for the more you get.[39] Some jurisdictions also allow explicit mention of some parameters of the defendant's financial condition, and this factor has been shown to serve as an anchor for award amounts.[40] As mentioned above, however, it is far from clear which financial parameter a particular jury will seize upon as an anchor.

Principle: Give them the right information. Most normative theories of decision-making and rational behavior agree on the principle that the more complete and the more relevant the information on which a decision is based, the better it will be.[41] The problem of vagueness in instructions and training can sometimes be solved by providing concrete examples from decisions in similar situations as a guide.[42]

Practice: Critical information is missing. The design of the jury trial system provides an abundance of information in certain areas, but fails to do so in other critical phases. Consistent with the principle of giving people the right information, the current punitive damage system places tremendous emphasis on the presentation of highly detailed and well-argued information about many aspects of a case. For example, the civil justice system gives particular attention to correct establishment of the facts of the situation that gave rise to the action. In many ways, the current regime does an excellent job on this point.

Given this careful attention to presenting and arguing evidence on some aspects of a case, it is remarkable that the system largely ignores other important aspects. The instructions about an appropriate dollar amount are vague, and jurors are explicitly prohibited from considering

---

[37] Sunstein, Kahneman & Schkade, *supra* note 7, at 2109–10; *See* W. Kip Viscusi, 39 HARV. J. ON LEGIS. 139, 160–61 (2002)

[38] *Cf. id.* (examining the effects of dollar anchors presented by plaintiff's attorneys on punitive damage awards). The defendant's counsel can also suggest a number, of course. However, since the position of the defendant is that punitive damages should not be awarded, this number is likely to be very small or zero, and therefore likely to be mostly ignored once the jury has moved on to the task of deciding on a specific non-zero amount of damages.

[39] Reid Hastie, David A. Schkade & John W. Payne, *Juror Judgments in Civil Cases: Effects of Plaintiff's Requests and Plaintiff's Identity on Punitive Damage Awards*, 23 LAW & HUM. BEHAV. 445 (1999); Gretchen B. Chapman & Brian H. Bornstein, *The More You Ask For, the More You Get: Anchoring in Personal Injury Verdicts*, 10 APPLIED COGNITIVE PSYCHOL. 519 (1996) (finding similar effects for plaintiffs' demands in the context of compensatory awards for personal injury).

[40] Kahneman, Schkade & Sunstein, *supra* note 3. Also recall the examples above of advertising budgets or total sales serving as anchors.

[41] *See, e.g.*, Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 207–09 (1968); PFEFFER, *supra* note 9, at 85–90.

[42] For limitations of this solution, see *infra* Part III.

punitive damage awards in other cases that might provide a context for assessing an amount of damages.[43] Jurors do appear to be able to assess the relative outrageousness of a defendant's behavior with reasonable consistency.[44] The difficulty of translating this judgment into a specific dollar amount is the major cause of unpredictability in punitive damage awards.[45]

Information about the magnitude of punitive damages awarded in similar cases might be particularly helpful to jurors. Looking for analogous situations as a guide and investigating their outcomes is an obvious and prudent step in making any significant decision. The prohibition against considering other cases or their outcomes robs jurors of critical points of reference that could help to ensure that similarly situated parties are indeed treated similarly across a range of cases.[46]

The recent Supreme Court opinion in *Cooper v. Leatherman*[47] emphasizes the importance of comparisons to other cases, even for experienced judges.[48] This decision compels future appellate courts to make explicit comparisons to other cases in a *de novo* review of the size of a punitive damage award.[49]

Principle: Give them incentives to perform well. If we reward people for the desired behaviors and outcomes, they will figure out a way to perform well.

---

[43] D. C. Barrett, *Propriety & Prejudicial Effect of Reference By Counsel in Civil Cases to Amount of Verdict in Similar Cases*, 15 A.L.R. 3d 1144 (1968).

[44] *See* Kahneman, Schkade & Sunstein, *supra* note 3, at 51–61.

[45] *See id.*; *see also* Sunstein, Kahneman & Schkade, *supra* note 7, at 2078–2103.

[46] Of course, this remedy would only address concerns about inconsistency across similar cases, and would not guarantee that any given award was correct. The problem here is that if the awards in comparison cases are themselves flawed, the approach might only achieve consistent injustice. *See* Cass R. Sunstein, Daniel Kahneman, David Schkade & Ilana Ritov, *Predictably Incoherent Judgments*, Parts V–VIII (forthcoming 2001).

[47] *See* Cooper Indus., Inc. v. Leatherman Tool Group, 121 S. Ct. 1678, 1687–88 (2001).

[48] *See id.* at 1687–88, stating:

> Differences in the institutional competence of trial judges and appellate judges are consistent with our conclusion. In *Gore*, we instructed courts evaluating a punitive damages award's consistency with due process to consider three criteria: (1) the degree or reprehensibility of the defendant's misconduct, (2) the disparity between the harm (or potential harm) suffered by the plaintiff and the punitive damages award, and *(3) the difference between the punitive damages awarded by the jury and the civil penalties authorized or imposed in comparable cases.* Only with respect to the first *Gore* inquiry do the district courts have a somewhat superior vantage over courts of appeals, and even then the advantage exists primarily with respect to issues turning on witness credibility and demeanor. Trial courts and appellate courts seem equally capable of analyzing the second factor. And the third *Gore* criterion, which calls for a broad legal comparison, seems more suited to the expertise of appellate courts. Considerations of institutional competence therefore fail to tip the balance in favor of deferential appellate review (emphasis added, citations and footnote omitted).

[49] *See id.* at 1685–88.

The ability of individuals to shape their behavior to fit the incentives they are given is an article of faith in most conceptions of rational human behavior.[50] Indeed, human beings often do find unanticipated and ingenious methods to achieve defined goals, even in the absence of obvious approaches. This ability to innovate is particularly strong when there is the opportunity to learn from repeated experiences with the same type of task.

This principle is perhaps most widely understood under the concept of "management by objectives" or the "bottom line" mentality, in which the goals for a person or subunit in an organization are specified, but the means for achieving these goals are not.[51] The person or subunit is then held accountable for achieving the goals within some specified period of time.[52] This approach would seem to work best with the principle of hiring the best people for a given job. In this way, the details of the task can be left more ambiguous, since the experience and creativity of the employees can be relied on to find a solution.

Practice: Jurors get neither rewards nor feedback based on their decisions. In practice, the role of jurors determining punitive damages is poorly suited to this principle. Despite having almost unlimited discretion in choosing an award amount, jurors determining punitive damages have neither performance-based incentives nor accountability for their decision. They are usually paid a flat daily amount for their service (anywhere from $5 to $50 per day depending on the state and the circumstances), which is not only small in magnitude but also does not depend in any way on the content of the decision they make.[53] After completing jury service, a juror is released to resume life as before, as if the trial had never occurred, and is accountable only to his or her own conscience (unless there was misconduct of some kind).[54] Further, they receive no feedback and or opportunity to learn from their mistakes, since being a punitive damage juror is typically a once-in-a-lifetime experience.[55] By

---

[50] *See, e.g.*, PFEFFER, *supra* note 9, at 213–17. *See also* WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 160–66, 184–85, 222–24 (1987); Becker, *supra* note 41, at 207–09;

[51] GEORGE S.ODIORNE, MANAGEMENT BY OBJECTIVES: A SYSTEM OF MANAGERIAL LEADERSHIP (1965); JOHN B. MINER & DONALD P. CRANE, HUMAN RESOURCE MANAGEMENT: THE STRATEGIC PERSPECTIVE 256–57 (1995); RICHARD M. STEERS & J. STEWART BLACK, ORGANIZATIONAL BEHAVIOR, 211–12 (5th ed. 1994).

[52] R. Rodgers, J. E. Hunter, *A Foundation of Good Management Practice in Government—Management by Objectives*, 52 PUB. ADMIN. REV. 27–39 (1992).

[53] *See* BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, *supra* note 12, Table 40 at 269–72; *see also* Strier, *supra* note 16, at 72–74.

[54] In *BMW*, 517 U.S. at 596 (Breyer J., concurring) ("one cannot expect to direct jurors like legislators through the ballot box; nor can one expect those jurors to interpret law like judges, who work within a discipline and hierarchical organization that normally promotes roughly uniform interpretation and application of the law").

[55] Priest, *supra* note 14, at 188–89.

design, therefore, jurors are always novices, at the beginning of the learning curve.

Principle: Provide a review process for exceptional cases. In principle, a system of review that assigns more experienced decision makers (e.g., managers or administrators) to review unusually difficult or novel cases is a sound principle of organization design.[56] In this type of system, job descriptions for roles lower in the hierarchy can be made simpler and more specific, since all possible contingencies need not be specified *ex ante*.[57] If all else fails, those few decisions that are unusually complex or otherwise questionable can be reviewed by more experienced people and modified or replaced. The success of this principle depends, however, on the existence of only a modest frequency of exceptions that arise for review.[58] Otherwise the hierarchy of the organization can be overloaded by the handling of exceptions, leaving insufficient time for other important activities such as planning, coordination, and strategic analysis.[59] In the context of the legal system, this overload means both delays and possibly a smaller number of cases heard or reviewed.

Practice: Judges must frequently correct jury awards. In practice, the rate of "exceptions" in punitive damage cases is staggering. Judges find it necessary to alter or reject *more than half* of jury punitive damage awards in product liability cases.[60] This is a startling statistic. When verdicts are appealed, as many as half of those awards are reduced or overturned, despite the apparent delegation of this task to juries.[61] Such frequent judicial intervention suggests that the legal system already implicitly recognizes that most juries struggle with this task. It is not unfair to say that this system has a high failure rate.[62]

Some observers see this frequent use of judicial review as a salutary feature of the punitive damages regime, as a sign that the system is working as intended.[63] There are, however, many consequences to leaning so heavily on this principle. With this basic philosophy one is constantly correcting mistakes after they are made, rather than preventing them in

---

[56] *See* JAMES G. MARCH & HERBERT A. SIMON, ORGANIZATIONS 137–71 (1958).

[57] *See id.* at 158–59.

[58] *See* JAY GALBRAITH, DESIGNING COMPLEX ORGANIZATIONS 11–15 (Edgar H. Schein et al. eds., 1973).

[59] *Id.*

[60] *See, e.g.*, Marc Galanter, *Real World Torts: An Antidote to Anecdote*, 55 MD. L. REV. 1093, 1115–16 (1996) (citing studies finding that jury punitive damage awards are reduced in fifty percent or more of product liability cases).

[61] A recent analysis by Eisenberg and Clermont reported in the *New York Times* found that nearly 40% of jury awards in injury and contract cases are reduced or overturned. *See* Glaberson, *supra* note 15.

[62] A similarly high rate of judicial intervention in criminal cases would probably be considered scandalous.

[63] *See, e.g.*, David Luban, *A Flawed Case Against Punitive Damages*, 87 GEO. L.J. 359, 362 (1998); Neil Vidmar, *Juries Don't Make Legal Decisions! And Other Problems: A Critique of Hastie et al. on Punitive Damages*, 23 LAW & HUM. BEHAV. 705, 711–12 (1999).

the first place—the opposite of good systems design. One cost is a contribution to case overload in the trial and appellate courts, i.e., the greater the frequency of surprising awards, the greater the frequency of appeals. To the extent that the court system is already overloaded, this only makes things worse. Another damaging consequence is the transaction cost of appeals. Appeals usually involve significant and sometimes severe delays, additional legal costs, and other opportunity costs. The possibility of protracted litigation and its associated uncertainty is one factor that contributes to the sometimes pathological risk aversion that is reflected in some out-of-court settlements.[64]

In view of the failure of the punitive damages task to adhere to these five task design principles, it is no wonder that juries struggle to find an appropriate amount of punitive damages. Incredibly, the current regime almost seems designed to encourage erratic awards.[65] It is therefore largely reactive, relying more heavily on catching and correcting errors after they are made, rather than on preventing the errors in the first place.

III. What Can Be Done?

Assessing punitive damages is an important legal and societal decision, one that deserves a well-designed system. The current system for deciding punitive damages falls short, and in some cases far short, of satisfying even simple task design principles. Jurors are compelled to perform a task—assessing a specific amount of punitive damages—that they cannot possibly be expected to perform well. Yet most jurors no doubt take their charge extremely seriously, and give it their best effort. If you ask them for a number, they will probably find a way to give you one. However, although necessity is the mother of invention, the civil justice system is not intended to be an opportunity for juror creativity. Indeed, the controversy over punitive damages has occurred despite jurors' valiant but doomed efforts to do the right thing. If their decisions are erratic, it is not the result of flawed jurors, but of a flawed design.

Can the task of determining an amount of punitive damages be modified so that it is sensible for jurors to perform it? Consider each of the design principles, and the type of reform that is suggested by the current gap between principle and practice.[66]

---

[64] *See generally* Kenneth R. MacCrimmon & Donald A. Wehrung, Taking Risks: The Management of Uncertainty (1986) (describing possible executive responses to large risks); Richard L. Manning, *Products Liability and Prescription Drug Prices in Canada and the United States*, 40 J.L. & Econ. 203, 210–34 (1997) (estimating the degree to which product liability costs in the United States contribute to the observed differences in pharmaceutical prices between the United States and Canada); Paul H. Rubin et al., BMW v. Gore: *Mitigating the Punitive Economics of Punitive Damages*, 5 Sup. Ct. Econ. Rev. 179, 192–96 (1997) (describing trends in and effects of punitive damage awards).

[65] *See Haslip*, 499 U.S. at 42–64 (O'Connor, J., dissenting).

[66] For a more detailed discussion of the reform possibilities discussed here (as well as

Can we get better people? One of the most important values embodied in the jury system is the participation of ordinary citizens in the administration of justice.[67] This participation enhances the transparency and perceived fairness of the system. To increase the knowledge and experience of jurors about the substance of a given case would require substantially greater selectivity, which could conflict with these values. In particular, if we select jurors with relevant skills and experience, some people would necessarily become more likely than others to serve as jurors.[68] It would probably be difficult to implement this selectivity in a way that was perceived as fair. Even if this selectivity did exist, the problem that few if any jurors would have ever had any experience deciding on punitive damages would remain, leaving that aspect of experience at a low level.

Can we give them better training? The biggest challenge to improving the training (i.e., jury instructions) is that even legal theorists cannot agree on a uniform theory of punitive damages.[69] Since there is no agreement about the appropriate amount, given a particular set of facts, it is not clear what modifications should be made to the current instructions. It is possible that better explanations of the meaning of key terms in the instructions (e.g., reprehensibility, subjective awareness, recklessness) could be provided. The most severe problem of assessing an amount of damages is the *translation* problem: how to translate a moral judgment about the level of punishment a defendant deserves into a specific dollar amount.[70] Without a theory regarding the appropriate amount, it is not clear how juries could be instructed about the process of translation. Even if a particular theory were adopted, it is not obvious even then that jurors would be both willing and able to perform the necessary calculations.[71]

Can we show them comparison cases? The most straightforward idea would be to allow juries to examine the punitive damage awards in

---

others), *see* CASS R. SUNSTEIN, REID HASTIE, JOHN W. PAYNE, DAVID A. SCHKADE & W. KIP VISCUSI, PUNITIVE DAMAGES: HOW JURIES DECIDE (forthcoming 2002); Sunstein, Kahneman & Schkade, *supra* note 7, at 2112–29; Sunstein, Kahneman, Schkade & Ilana Ritov, *supra* note 46, at Parts V–VIII.

[67] Landsman, *supra* note 18, at 579–81, 588–89.

[68] See Strier, *supra* note 16, at 72–79. *See also* Lilly, *supra* note 16, at 72–79, 83–89.

[69] For examples of the theoretical debate about the appropriate level of punitive damages, see Landes & Posner, *supra* note 50, at 160–66, 184–85, 222–24 (1987) (advocating the economic theory of optimal deterrence); Galanter & Luban, *supra* note 6, at 1404–51 (arguing that punishment, rather than deterrence, is the more important function of punitive damages).

[70] *See* Sunstein, Kahneman, Schkade & Ritov, *supra* note 46, Part III.B. *See also* Kahneman, Schkade & Sunstein, *supra* note 3, at 51–55.

[71] *See* Viscusi, *supra* note 37, at 163. *See also* Cass R. Sunstein, David Schkade & Daniel Kahneman, *Do People Want Optimal Deterrence?*, 29 J. LEGAL STUD. 237, 249 (2000) (considering jurors' willingness to follow the provided instructions); W. Kip Viscusi, *The Challenge of Punitive Damages Mathematics,* 31 J. LEGAL STUD. 351, 351–82 (2001) (considering jurors' ability to follow the provided instructions).

other cases during their own deliberations. Closer inspection shows this apparently simple move to be fraught with difficulties.

The first problem involves cognitive overload. Which specific cases should be shown to the jury? How many? In what detail? Obviously, it would be absurd to provide jurors with all similar cases. Any set of comparison cases would be arbitrary.

The second problem involves the risk of manipulation. It would be easy to select a comparison set that increased dollar awards, or alternatively to select a set that decreased dollar awards.[72] A lack of a comparative context is indeed a problem, but allowing the jury to make explicit comparisons is probably not a realistic response.

Can we give them performance-based incentives? As a practical matter, jury service is an unpaid duty in which jurors are compensated mainly by their pride in service. The token payments jurors do receive are both small and unconnected to the quality of their decisions.[73] Connecting jurors' payments in any way to their decision goes strongly against the grain. Suppose that in an attempt to establish some limited accountability a legislator proposed that when a punitive damage award is altered or rejected by a trial or appellate judge (which as we saw earlier happens over half the time), the jurors would be recalled to justify their decision. If their explanation was inadequate, their pay would be reduced. Even this simple attempt at establishing the most minimal accountability seems completely impractical, and would probably be considered outrageous by many. It is not clear even in principle how one could develop a proper incentive scheme for jurors so long as jury service remains a civic duty and not a job.

Can we strengthen judicial review? The most plausible and least disruptive alternative is to reallocate some or all of the jury's task to judges. Judges have the experience of having made decisions in many cases, are well trained in the law, have access to a large number of documented cases for comparison, and can be held at least somewhat accountable for their decisions (thus to some degree aligning their rewards with their decisions).[74] While far from perfect, giving judges a greater role fully or partially addresses current shortfalls against the first four task design principles (and therefore also the fifth). Another virtue of this direction is that bench trials take less time and cost less, on average, than jury trials.[75] Due to the already very high rate of judicial intervention in altering jury punitive damage awards, this does not seem a radical suggestion.

---

[72] *See* Sunstein, Kahneman, Schkade & Ritov, *supra* note 46, at Part V.B.1.

[73] *See, e.g.*, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, *supra* note 12, Table 40 at 269–72.

[74] *See* Strier, *supra* note 16, at 73.

[75] *See* Lilly *supra* note 16, at 57–59 (citing evidence that jury trials take two to three times as long and cost up to ten times more than bench trials).

The most straightforward version of this proposal is to continue to allow juries to decide questions of liability for punitive damages, just as they do questions of criminal liability. As judges in criminal trials make the determination of punishment (subject to guidelines), judges in civil trials could have an exclusive or at least a larger role in selecting the level of punitive damages.[76] Giving judges the exclusive responsibility for assessing the amount of punitive damages, while still letting juries decide whether punitive damages should be awarded, would address most of the concerns raised above, while substantially retaining public participation. However, there is a material loss in participation, due to the removal of the opportunity for direct expression of the *degree* of community disapproval by the jury through the amount of damages (however unreliable the expression of this sentiment in dollars may be).[77]

At the very least, judges should be encouraged to review punitive awards carefully to make sure that they are sensible in light of the various goals of the legal system, including consistency with other similar cases. A minimal reform would then be to give clear directions to judges to scrutinize jury awards by making sure that they are not out of line with awards made in other similar cases.[78]

Of course, judges themselves are not immune to the problems faced by jurors in this challenging task, even if the problems are reduced when judges are involved.[79] In the criminal context, sentencing guidelines were introduced in the 1980s to limit judicial discretion in sentencing. In both the civil and the criminal contexts, the problem is that different judges might reach different conclusions in similar cases, thus producing arbitrariness.[80] In the context of punitive damages, because of the difficulty and ambiguity of the task of determining a specific amount, and despite their experience and training, it is not only possible but likely that different judges will arrive at conflicting damage awards. Hence, there is good reason for guidelines and constraints on judges in the punitive damages context. Under *Leatherman*, it is now clear that the appellate courts are required to compare a punitive damage award to those in other cases, although the exact form and method of this comparison remains to be

---

[76] SUNSTEIN, HASTIE, PAYNE, SCHKADE & VISCUSI, *supra* note 66, chapter 13 (forthcoming).

[77] *See* Kahneman, Schkade & Sunstein, *supra* note 3, at 49–78; Sunstein, Kahneman & Schkade, *supra* note 45, at 2127–28.

[78] Indeed, this is the very direction that the Supreme Court has given to appeals courts. *See Leatherman*, 149 L.Ed.2d at 684–91.

[79] *See* Chris Guthrie, Jeffrey J. Rachlinski, & Andrew J. Wistrich, *Inside The Judicial Mind*, 86 CORNELL L. REV. 777 (2001); W. Kip Viscusi, *How Do Judges Think About Risk*, 1 AM. L. & ECON. REV. 26, 26–61 (1999); W. Kip Viscusi, *Jurors, Judges, and the Mistreatment of Risk by the Courts*, 30 J. LEGAL STUD. 107, 107–36 (2001).

[80] Indeed, substantial differences between judges and jurisdictions in average sentences for the same crime were one of the concerns that motivated the establishment of the federal criminal sentencing guidelines. *See Conference on the Federal Sentencing Guidelines: Summary of Proceedings* 101 YALE L.J. 2053, 2053 (1992).

established.[81] For example, trial judges might also be required to conduct their own determination of punitive damages in order to provide a point of comparison in evaluating the jury's decision.

IV. Conclusion

If jury punitive damage awards surprise us and seem unpredictable, it is due more to the situation that jurors find themselves in than to the characteristics of the jurors themselves. Several features of the task of assessing an amount of punitive damages seem almost designed to produce erratic awards. Unfortunately, it is not clear that these flaws can be corrected while retaining the full participation of the jury. The discussion here has focused on reforms related to the division of labor between judges and juries because of the close connection of this avenue to the design principles presented above. There are, however, several other interesting possible directions for reform, including damage schedules and administrative translation of jury judgments into dollars, which have been detailed previously.[82] Whatever the reform, we must take care to avoid requiring people to perform a task that they cannot possibly be expected to perform well.

---

[81] *See Leatherman*, 121 S. Ct. at 1687–88.
[82] *See supra* Part III.

TABLE I. PRINCIPLE AND PRACTICE IN ASSESSING PUNITIVE DAMAGES

PRINCIPLE	THEORY	PRACTICE
1. Hire the best people	Get the best decision makers you can find, people who have extensive knowledge, training, and experience in making this type of decision.	Punitive damages are set by low paid, randomly selected draftees, who after the filtering of the *voir dire* process have little relevant knowledge, training or experience.
2. Give them good training	If we can't get people with skill and experience, we can give those we can get specific instructions and clear criteria, and if they follow this guidance they will still perform well.	Typical punitive damage instructions are vague principles, rather than concrete actions or specific criteria. There is little guidance about what an appropriate amount would be. Jurors often ignore these instructions and substitute their own criteria.
3. Give them good information	The problem of vagueness can sometimes be solved by providing a context of concrete examples from decisions in similar situations.	Critical information is missing. Punitive damage jurors are explicitly prohibited from considering other cases. This makes determining an appropriate range almost impossible.
4. Give them incentives to perform well	If we reward people for the desired behaviors and outcomes, they will figure out a way to perform well.	Being a punitive damage juror is a one-shot deal, with neither performance-based incentives nor accountability for the decision made.
5. Provide a review process for exceptional cases	If all else fails, those few decisions that are unusually difficult ("exceptions") can be reviewed by more experienced people.	Apparently, most cases are exceptions. More than half of jury punitive damage awards are reduced or thrown out by trial or appellate judges.

ARTICLE

PUNITIVE DAMAGES: HOW JURORS FAIL TO PROMOTE EFFICIENCY

W. KIP VISCUSI*

Evidence of corporate risk-cost balancing often leads to inefficient punitive damages awards, suggesting that jurors fail to base their decision-making on principles of economic efficiency. In this Article, Professor Viscusi presents the results of two experiments regarding jury behavior and punitive damages. In the first experiment, Professor Viscusi found that mock jurors punish companies for balancing risk against cost, although award levels vary depending on how the economic analysis is presented at trial. The results of the second experiment suggested that mock jurors are unwilling or unable to follow a set of model jury instructions designed to generate efficient damages awards. Professor Viscusi concludes that neither risk-cost analysis nor this particular set of instructions can encourage jurors to behave efficiently. As a result, damages awards may create undesirable incentives for companies making choices about safety.

I. INTRODUCTION

The challenges associated with determining punitive damages loom among the greatest difficulties juries face in our civil justice system.[1] It is well known that juries may fail in some respects. The more interesting issue is whether jury performance can be improved, or whether jury failings can be eliminated altogether.

What do we mean by juror performance being a success or failure? Three different reference points will be used in this Article. The first and most rigorous reference point is that punitive damages should be efficiency-enhancing. Jurors should set levels of punitive damages so as to create efficient incentives for deterring reckless behavior. Firms consequently should have financial incentives to take the appropriate degree of care and manufacture sufficiently safe products.

---

* John F. Cogan, Jr. Professor of Law and Economics, Harvard Law School. This research was supported by the Sheldon Seevak Research Fund; the Olin Center for Law, Economics and Business at Harvard Law School; and a grant to the author from the Exxon Corporation.

[1] For an analysis of jurors' inability to translate their outrage into dollar awards, see Cass R. Sunstein, Daniel Kahneman & David Schkade, *Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)*, 107 YALE L.J. 2071 (1998). Punitive damages' failure to produce beneficial safety effects is the focus of W. Kip Viscusi, *Why There Is No Defense of Punitive Damages*, 87 GEO. L.J. 381 (1998). *See also* W. Kip Viscusi, *The Social Costs of Punitive Damages Against Corporations in Environmental and Safety Torts*, 87 GEO. L.J. 285 (1998).

A second test is more limited. Suppose one does not accept economic efficiency as the goal of punitive damages. Nevertheless, firms should be punished less for greater efforts to promote safety. The more a firm invests in safety, the more unlikely it is that the firm should be judged reckless for displaying a conscious disregard for safety. Holding constant other aspects of the firm's behavior, as the level of investment in safety increases, the frequency and level of punitive damages should fall.

The third and final test is still more limited. Do jurors properly implement formal instructions when setting punitive damages levels? Wholly apart from broader law and economics norms, do jurors adhere to specific instructions from the judge telling them how to make efficient awards?[2]

Unfortunately, jury damages awards tend to discourage safe corporate behavior by punishing careful decision-making, even when jurors have been explicitly instructed to act efficiently. This Article examines two efforts directed at improving jury performance in balancing risk and cost and in adhering to jury instructions. The first set of results pertains to jury judgments of recklessness when a defendant company performed a risk-cost analysis.[3] Mock jurors considered case scenarios and were asked to assess whether the company's behavior was reckless and warranted punitive damages. Ideally, companies should undertake systematic risk analyses to achieve the appropriate balance of risk and cost in their decision-making.[4] In fact, corporate risk analyses have been associated with many of the most prominent punitive damages awards, particularly with respect to the automobile industry.[5] This suggests that jurors interpret risk-cost analyses as worthy of punishment rather than as praiseworthy efforts to improve safety. Part III of this Article explores how different case contexts affect the likelihood and level of punitive damages. Building on results from an earlier series of experiments, I found that jurors fail to make efficient punitive damages awards and are vulnerable to dollar values that are suggested as anchor values at trial.[6]

---

[2] Whether punitive damages should be set by a judge or eliminated for many classes of cases involves judgments beyond the scope of this paper.

[3] In the usual policy analysis case, there is a benefit-cost comparison. Preventive steps are taken if the value of the benefit exceeds the costs of attaining that benefit. For policy choices involving risk, the benefit is risk reduction, leading to a tradeoff between risk reduction and cost, or a risk-cost tradeoff.

[4] The approach I take here is the standard policy analysis perspective. *See* W. KIP VISCUSI, FATAL TRADEOFFS: PUBLIC & PRIVATE RESPONSIBILITIES FOR RISK 4–6 (1992) [hereinafter VISCUSI, FATAL TRADEOFFS]. Judge Frank Easterbrook describes the desirability of corporations using their expertise to make safety decisions in *Carroll v. Otis Elevator Co.*, 896 F.2d 210, 215 (7th Cir. 1990).

[5] *See, e.g.*, Jimenez v. Chrysler Corp., 74 F. Supp. 2d 548 (D.S.C. 1999), *vacated by* 2001 U.S. App. LEXIS 22562 (5th Cir. Oct. 19, 2001); General Motors Corp. v. Moseley, 447 S.E.2d 302, 311 (Ga. Ct. App. 1994).

[6] The results of the initial series of experiments were published in W. Kip Viscusi, *Corporate Risk Analysis: A Reckless Act?*, 52 STAN. L. REV. 547, 553 (2000) [hereinafter Viscusi, *Corporate Risk Analysis*]. Additional analysis of the results may be found in *Ju-*

Part IV presents the results of a study on the effectiveness of jury instructions developed by Polinsky and Shavell in an effort to ground the determination of punitive damages in sound law and economics principles.[7] Typical punitive damages instructions ask jurors to determine whether the defendant's conduct was reckless, i.e., whether the defendant was conscious of grave danger, whether that risk eventuated, whether the defendant disregarded the risk, and whether the defendant's conduct was a gross deviation from the level of care an ordinary person would select.[8] If these criteria are met, the jury may award punitive damages to punish and deter the defendant. Armed with the Polinsky-Shavell instructions, however, the determination of punitive damages should be a straightforward mathematical exercise involving little more than simple multiplication and addition.[9] By reducing the task of setting punitive damages to the implementation of a formula, the Polinsky-Shavell approach ideally should eliminate the random element of punitive damages and ground them in important principles of efficiency.[10] In practice, the instructions do not appear to improve the efficiency of damages awards.

My assessment of juror reactions to corporate risk analyses and to the Polinsky-Shavell punitive damages instructions was based on controlled experimental results involving hundreds of jury-eligible citizens. The experimental findings do not bode well for the possibility of remedying perceived inadequacies in jury performance. Jurors do not seem receptive to risk-cost balancing and also fail to implement the guidance offered by the Polinsky-Shavell formulas. Jurors are vulnerable to suggestions made at trial regarding appropriate awards, even when they have been directed to use a different method for determining damages. Furthermore, the damages awards made in the studies penalized companies that made efficient decisions about risk. It appears that instead of deterring risky conduct, damages often discourage companies from taking efficient steps to make their operations safer.

---

rors, Judges, and the Mistreatment of Risk by the Courts, 30 J. LEGAL STUD. 107, 119–27 (2001) [hereinafter Viscusi, *Jurors*].

[7] *See* A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 957–62 (1998). Results from this study also are analyzed in W. Kip Viscusi, *The Challenge of Punitive Damages Mathematics*, 30 J. LEGAL STUD. 313 (2001) [hereinafter Viscusi, *Challenge of Punitive Damages*]

[8] *See* Jardel Co., Inc. v. Hughes, 523 A.2d 518, 527–31 (Del. 1987).

[9] See Polinsky & Shavell, *supra* note 7, at 960–62 for their formula for setting punitive damages for firms.

[10] While Polinsky and Shavell stress the importance of the formula in providing "guidance" to jurors with respect to punitive damages levels, as a practical matter the variability of awards should be reduced as well. *See id.* at 954–56. For example, the calculation of the deterrence value of punitive damages has only one correct answer for any particular case. If jurors follow the formula, then their damages awards should not vary from this figure. *See id.* at 960–61.

II. THE PRACTICAL CONSEQUENCES OF CORPORATE RISK ANALYSES[11]

In order for businesses to make efficient decisions about safety, risk-cost balancing must be an explicit concern of corporate risk decision-making. Such balancing lies at the heart of standard negligence tests, particularly those that are framed within the context of a law and economics approach, such as the Learned Hand rule.[12] Risk-cost tradeoffs are central to the commonly used risk-utility test for assessing liability, which balances a product's usefulness and safety characteristics to determine whether the company has attained the correct balance of cost and risk.[13]

Similarly, risk-cost tradeoffs play a central role in government regulatory policy in the risk and environmental area, because a regulation's stringency is reflected in the cost expended by regulated companies per life saved.[14] For any given regulatory policy, it is usually the case that the costs of the regulation become increasingly greater as the regulation is tightened.[15] These tradeoffs are incurred as the government seeks to induce regulated industries to achieve desired levels of safety.[16] It should be noted, however, that the decision-making processes of regulators and juries are not identical. The government sets regulatory levels and businesses set care levels in anticipation of a certain number of future accidents. At the time the lifesaving decisions are made, the potential victims who would be the beneficiaries of additional safety precautions generally are not known. In contrast, juries consider accident cases in which there is an identified victim. Jurors may find it difficult to evaluate the business's conduct from an *ex ante* perspective once there is a specific, human victim in addition to a statistical probability of harm.[17]

---

[11] The discussion in Part II draws extensively on Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 548–52, 567–78. For additional analysis, see Viscusi, *Jurors*, *supra* note 6, at 117–18.

[12] *See* A. MITCHELL POLINSKY, AN INTRODUCTION TO LAW AND ECONOMICS 41–42 (2d ed. 1989); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 179–83 (5th ed. 1998). The Learned Hand formula parallels the risk-cost tradeoff in cost-benefit analysis. A firm is negligent if the cost of the precaution is less than the probability of an accident multiplied by the size of the loss incurred when the accident does occur. *See* POSNER at 180.

[13] *See* John W. Wade, *On the Nature of Strict Tort Liability for Products*, 44 MISS. L.J. 825, 837–38 (1973); W. KIP VISCUSI, REFORMING PRODUCTS LIABILITY 70–81 (1991). I discuss this issue in greater detail in Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 549.

[14] *See* VISCUSI, FATAL TRADEOFFS, *supra* note 4, at 4–6.

[15] *See* W. KIP VISCUSI, RISK BY CHOICE: REGULATING HEALTH AND SAFETY IN THE WORKPLACE 114–15 (1983). Justice Breyer refers to the escalation of costs as regulations are tightened as "the problem of the last 10 percent." In other words, such stringent regulations impose "high costs without achieving additional safety benefits." *See* STEPHEN BREYER, BREAKING THE VICIOUS CIRCLE: TOWARD EFFECTIVE RISK REGULATION 10–19 (1993).

[16] *See* BREYER, *supra* note 15, at 9–10. For a more detailed discussion, see Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 561–62.

[17] *See* Carroll v. Otis Elevator Co., 896 F.2d 210, 215–16 (7th Cir. 1990).

Because of the central role played by risk-cost tradeoffs in attaining desired levels of care, a recent preliminary draft of the Restatement (Third) of Torts explicitly recognized the importance of undertaking such an analysis as well as the fact that doing an analysis in and of itself should not be a basis for finding negligence:

> Under § 4, negligence is defined in terms of the failure to exercise reasonable care, and reasonable care is explained primarily in terms of the balance between the magnitude of the foreseeable risk and the burden of precautions that can eliminate the risk. If the burden is greater than the risk, the actor who declines to adopt that precaution is not negligent. But if the magnitude of the risk is somewhat greater than the burden, the actor is negligent for failing to adopt the precaution.
>
> From this evaluation, two points follow that relate to the meaning of recklessness. The first point is a negative one: the fact that the actor, because of the burden entailed by a particular precaution, has made a deliberate choice to omit a precaution and hence to tolerate a risk by no means signifies that the person has behaved recklessly. Indeed, the fact that such a choice has been made does not even show that the actor has behaved negligently. Rather, the actor is negligent only for making an unwise choice. In a sense, the very objective of negligence law is to encourage actors to acknowledge and confront such choices, and to render these choices wisely rather than unwisely.[18]

The practical experience with corporate risk analysis has not followed the Restatement guidelines, and plaintiffs often use the fact that the defendant conducted a risk analysis to infer that behavior was negligent or reckless. Indeed, as the discussion below explains, the analysis in and of itself can become a source of controversy and lead jurors to impose punitive sanctions on the company.[19]

In some cases, these corporate risk analyses occur after the fact, as the company attempts to determine the causes of a major accident. After airplane crashes, for example, companies should make a frank and critical assessment of the causes so as to prevent such accidents in the future.[20] Because jurors respond to these analyses by awarding higher dam-

---

[18] RESTATEMENT (THIRD) OF TORTS § 2, cmt. d (Discussion Draft 1999).

[19] Perhaps the most noteworthy instance was a General Motors case involving burn victims in a rear-ended Chevrolet Malibu, which will be discussed in detail below. *See infra* notes 31–36 and accompanying text.

[20] For a discussion of the tensions involved with post-crash investigations, see Donald S. Skupsky, *Legal Requirements for Records Prepared for Internal Investigations and Audits*, REC. MGMT. Q., Apr. 1992, at 34–36. I discuss the same issue in greater detail in Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 567–68.

ages, an investigation may jeopardize the company's prospects in court when the families of the accident victims seek compensation.[21] Defendants also may be concerned that a truly critical report could trigger punitive damages.[22] As a result, the frankness and thoroughness of the report could be compromised if the company were forced to release its own assessment of the accident.

Risk analysis following airline crashes was a central issue in the legal battle involving the 1979 American Airlines DC-10 crash near O'Hare Airport.[23] American Airlines had undertaken a detailed post-crash investigation, but later destroyed the report, claiming that it was covered by attorney-client privilege.[24] The court suggested that the failure to release the report did not necessarily imply that the report itself would have been damaging.[25] Nonetheless, one can speculate that American Airlines' decision to destroy the report may have been influenced by fear that jurors would view the analysis in a negative light.

An airline must make a difficult decision after a plane crash.[26] If it chooses not to undertake an accident investigation, surely that would be a sign of recklessness. The accident gave the company knowledge of a risk, and the company has a duty to investigate and determine how to reduce that risk in the future.[27] Undertaking some kind of post-accident assessment certainly would appear to be the more desirable course for preventing future harm.[28] Nonetheless, when the company realizes the punitive damages consequences of making such an analysis, it may choose to either forgo a thorough investigation or keep any resulting report secret.

In situations in which the corporate risk analysis occurs before the accident, the very act of undertaking the analysis often becomes a negative feature rather than a positive one in jury determinations of punitive damages.[29] Suppose the company makes a detailed risk assessment and then proceeds with a level of safety that is efficient, but that does not result in zero risk. Jurors often view the risk analysis as a negative aspect of corporate behavior, reasoning that the company knew of a way to make the product safer, but chose instead to endanger its customers.[30]

---

[21] *See* Skupsky, *supra* note 20, at 34–36.

[22] *See id.*

[23] *See In re* Air Crash Disaster near Chicago, Illinois on May 25, 1979, 90 F.R.D. 613 (N.D. Ill. 1981).

[24] *See id.*

[25] *See id.* at 621.

[26] This discussion draws extensively on Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 567.

[27] *See id.* Following any accident, one can analyze whether human or mechanical failure were contributory factors and whether changes in either would have affected the likelihood of an accident.

[28] *See* Skupsky, *supra* note 20, at 34.

[29] The cases discussed below will document this relationship. See *infra* notes 31–41 and accompanying text.

[30] *Id.*

A noteworthy case was that brought against General Motors ("G.M.") in Los Angeles, in which a Chevrolet Malibu was rear-ended on Christmas Eve, 1993. The driver, her four children, and a friend of the family were riding in the car, and all were seriously burned as a result of the accident.[31] During the trial, a memo by G.M. engineer Edward Ivey played a critical role leading to the $4.8 billion punitive damages award.[32] That memo analyzed the costs of fuel-fed, fire-related fatalities in cars and the costs of preventing these deaths.[33] Ivey valued fatalities at $200,000 each, which is the same value that Ford used in its analysis of risks associated with the Ford Pinto.[34] The Pinto analysis had used a number that paralleled the level of compensatory damages awards in fatality cases, which can be traced to the present value of lost earnings due to the accident.[35] This value was consistent with punitive damages awards in product-related fatality cases during that time period.[36]

Whether such an amount would be an appropriate yardstick to use when valuing life from the standpoint of preventing deaths in automobile crashes is more problematic. Two factors should be considered: the value from the standpoint of prevention, and the value from the standpoint of compensation.[37] In terms of preventing accidents, the appropriate economic measure is the value of a statistical life, or the risk-money tradeoff involving small risks.[38] For compensation purposes, looking at the income loss may be a reasonable measure of damages, but it will understate how much it is worth to prevent the death. People value their lives at more than their income level and they generally are not willing to accept certain death in return for compensation for future income loss.[39] Ivey's memo did not pursue these issues, but did conclude on a cautionary note:

> This analysis indicates that for G.M. it would be worth approximately $2.20 per new model auto to prevent a fuel fed fire

---

[31] For a description of this case, see Andrew Pollack, *Paper Trail Haunts GM After it Loses Injury Suit: An Old Memo Hinted at the Price of Safety*, N.Y. TIMES, July 12, 1999, at A12; Andrew Pollack, *$4.9 Billion Jury Verdict in GM Fuel Tank Case: Penalty Highlights Cracks in Legal System*, N.Y. TIMES, July 10, 1999, at A8 [hereinafter Pollack, *Jury Verdict*]; Ann W. O'Neill et al., *GM Ordered to Pay $4.9 Billion in Crash Verdict Liability*, L.A. TIMES, July 10, 1999, at A1.

[32] See Pollack, *Jury Verdict*, *supra* note 31, at A8.

[33] Memorandum from E. C. Ivey, to General Motors, Value Analysis of Auto Fuel Fed Fire Related Fatalities (June 29, 1973). Also see my earlier discussion of the Ivey memorandum in Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 574–76.

[34] *See* VISCUSI, REFORMING PRODUCTS LIABILITY, *supra* note 13, at 111. For further discussion of the Pinto case, see *infra* notes 44–51 and accompanying text.

[35] *See* VISCUSI, REFORMING PRODUCTS LIABILITY, *supra* note 13, at 111.

[36] *See id.*

[37] *See* W. Kip Viscusi, *Misuses and Proper Uses of Hedonic Values of Life*, 13(2) J. FORENSIC ECON. 111 (2000) [hereinafter Viscusi, *Misuses*].

[38] *See id.* at 112.

[39] *See* Viscusi, FATAL TRADEOFFS, *supra* note 4, at 39–40.

> in *all* accidents . . . . This analysis must be tempered with two thoughts. First, it is really impossible to put a value on human life. This analysis tried to do so in an objective manner but a human fatality is really beyond value, subjectively. Secondly, it is impossible to design an automobile where fuel fed fires can be prevented in all accidents unless the automobile has a non-flammable fuel.[40]

Was undertaking such an engineering analysis viewed as an honest attempt at risk-cost balancing? The attorney representing the plaintiffs argued just the opposite, claiming that this analysis showed that G.M. was "caught red-handed."[41] The company was aware of the risk and the potential for reducing it, yet proceeded with the original design. After the trial, one of the lawyers for the plaintiff observed that the large punitive damages award was, in effect, a critique of this approach to examining risk-cost tradeoffs: "The jurors wanted to send a message to General Motors that human life is more important than profits."[42] Jurors likewise highlighted the corporate risk analysis as significant in the decision-making process: "Jurors told reporters that they felt the company had valued life too lightly. 'We're just like numbers I feel, to them,' one juror, Carl Vangelisti, told Reuters. 'Statistics. That's something that is wrong.'"[43] These comments suggest that the large award reflected the jurors' belief that the very act of performing a risk-cost analysis merited punitive damages.

There are several ways for jurors to infer from a risk-cost analysis that punitive damages should be awarded. By conducting an analysis, the company demonstrates awareness of the risk. If it chooses not to adopt the safety measure, it will have demonstrated willingness to proceed with a design knowing that it would cost lives. The very act of undertaking an analysis may appear cold-hearted, and making an explicit tradeoff of lives for money may appear to be a gross deviation from proper care. At the time of the analysis, the tradeoff is between expensive across-the-board safety measures and a small expected loss to any particular individual. After an accident, the comparison is between a modest safety investment that would have prevented the specific fatality and an identifiable life, a trade-off that jurors may find reckless.

---

[40] Ivey, *supra* note 33 (emphasis added). Ivey's suggestion that it is impossible to obtain zero risk may run counter to jurors' intuition. From an efficiency perspective, however, the relevant issue is not whether the risk of the current design is greater than zero, but whether the risk of the current design is desirable when compared against the costs of risk reduction.

[41] *Jury Awards $4.9 Billion to Crash Victims, Finds GM Bargained Away Passenger Safety*, PROD. SAFETY & LIAB. REP., July 16, 1999, at 721.

[42] Pollack, *Jury Verdict*, *supra* note 31, at A8.

[43] *Id.*

In many respects, the experience with the Ivey analysis paralleled the earlier press treatment of the Ford analysis relating to the Pinto.[44] *Grimshaw v. Ford Motor Company*[45] focused on the safety aspects of the design of the Pinto and, in particular, the placement of the gas tank in the rear of the vehicle.[46] What was particularly noteworthy about *Grimshaw*, which led to a $125 million punitive damages award that was subsequently reduced to $3.5 million, was the role of corporate risk analysis in the media coverage of the case.[47] *Mother Jones* magazine published an article documenting that a Ford engineer had done a safety analysis on the Pinto assessing the benefits and costs of preventing fire-related deaths.[48] This article was released at a press conference featuring Ralph Nader and led to the awarding of a Pulitzer Prize to the magazine.[49] Ford's analysis was construed by *Mother Jones* as pertaining to the Pinto rear impact risks.[50] Although Ford was criticized for having analyzed the risks associated with rear-end impact fires, the document at issue was in fact an economic analysis prepared for the National Highway Traffic Safety Administration in opposition to a proposed regulation regarding fires associated with rollover risks.[51]

Similar issues arose with respect to the Chrysler minivan case, in which a six year-old, Sergio Jimenez, was thrown from the vehicle after his mother ran a red light and the minivan was struck on its side.[52] This case, which led to a $250 million punitive damages award against the Chrysler Corporation, was noteworthy in that the company had undertaken a post-accident analysis of the cost of fixing the allegedly defective rear-door latch that opened when the minivan rolled over after being hit

---

[44] *See* Mark Dowie, *Pinto Madness*, MOTHER JONES, Sept.-Oct. 1977, at 18.

[45] 174 Cal. Rptr. 348 (Cal. Ct. App. 1981).

[46] This case also receives extensive analysis in Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 568–70.

[47] *See* Gary T. Schwartz, *The Myth of the Ford Pinto Case*, 43 RUTGERS L. REV. 1013, 1020–21 (1991).

[48] *See* Dowie, *supra* note 44, at 18.

[49] *See* STUART M. SPEISER, LAWSUIT 357 (1980).

[50] *See* Dowie, *supra* note 44, at 24.

[51] *See* Schwartz, *supra* note 47, at 1020–21 & n.21. This mismatch between the *Mother Jones* report and *Grimshaw* is a central theme in Schwartz's article. *See also* BRENT FISSE & JOHN BRAITHWAITE, THE IMPACT OF PUBLICITY ON CORPORATE OFFENDERS 54 (1983), observing:

> In the absence of an offense defined in terms of manufacturing an unjustifiably dangerous product, questions of acceptable risk of the kind raised by the Pinto Papers will rarely be the central subject of inquiry in the context of corporate offenses against the person. This is unsatisfactory, not only because of the danger of a serious underlying risk being concealed from society, but also because it may do more harm than good not to face up to the need for studies of the costs of improving product safety in matters such as that for which Ford was pilloried.

[52] *See* Jimenez v. Chrysler Corp., 74 F. Supp. 2d 548 (D.S.C. 1999), *vacated by* 2001 U.S. App. LEXIS 22562 (5th Cir. Oct. 19, 2001). *See also* Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 551.

on the side.[53] The analysis concluded that the design did not pose a significant risk and that it would have cost $100,000 in fixed retooling costs as well as an additional $0.50 per vehicle to change the latch design.[54] After the trial, the plaintiff's attorney observed that "Chrysler officials at the highest level cold-bloodedly calculated that acknowledging the problem and fixing it would be more expensive, in terms of bad publicity and lost sales, than concealing the defect and litigating the wrongful death suits that inevitably would result."[55]

The 1984 case of *Ford Motor Co. v. Stubblefield*[56] involved a Ford Mustang II that caught fire after a rear impact.[57] The fire killed Terri Stubblefield, the passenger riding in the rear seat of the car. The jury awarded $8 million in punitive damages. The plaintiff's argument referred to a risk analysis undertaken by Ford and criticized the company for engaging in "safety science management":[58]

> The evidence here was sufficient to authorize the jury to find that the sum of $8 million was an amount necessary to deter Ford from repeating its conduct; that is, its conscious decision to defer implementation of safety devices in order to protect its profits. One internal memo estimated that "the total financial effect of the Fuel System Integrity program [would] reduce Company profits over the 1973–1976 cycle by $(109) million," and recommended that Ford "defer adoption of the [safety measures] on all affected cars until 1976 to realize a design cost savings of $20.9 million compared to 1974." Another Ford document referred to a $2 million cost differential as "marginal."[59]

The difficulties encountered by Ford with respect to its engineering risk analyses have not been limited to rear impact cases. In *Miles v. Ford Motor Co.*,[60] the company came under fire for its risk analysis done with respect to tension eliminator spools for lap belts.[61] In this particular case, the passenger, Willie Miles, leaned over to pick up some trash from the floor of the car, causing the shoulder harness to spool out and leaving a

---

[53] *See* Jimenez, 74 F. Supp. 2d at 560.

[54] *See* Donald C. Dilworth, *Jurors Punish Chrysler for Hiding Deadly Defect*, TRIAL, Feb. 1998, at 14, 16.

[55] *Id.* at 14.

[56] 319 S.E.2d 470 (Ga. Ct. App. 1984).

[57] For additional analysis of this case, see Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 572–73.

[58] 319 S.E.2d at 475.

[59] *Id.* at 481.

[60] 922 S.W.2d 572 (Tex. App. 1996), *remanded for procedural errors to* Ford Motor Co. v. Miles, 967 S.W.2d 377 (Tex. 1998).

[61] This case also is discussed in Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 572–73.

slack.[62] The tension eliminator spool did not rewind the belt.[63] The car was involved in a collision, and Miles slid through his lapbelt, catching his head and sustaining spinal injuries.[64]

Once again, the fact that Ford had conducted an analysis of this class of issues proved to be consequential in the arguments at trial.[65] Indeed, not only had Ford explored the issue, but it had undertaken an economic cost-benefit analysis.[66] Although such an analysis is necessary to achieve an efficient result, Ford's caution was presented in a negative light in court:

> Syson [the plaintiff's accident reconstruction expert] testified that he was familiar, during the relevant time period, with the corporate policies of Ford Motor Company as they related to potentially defective products. Syson testified that when Ford identified what it believed was a defective product it would first run a "cost benefit" analysis to see what the cost would be to fix or repair the defect. Next, Ford would assign arbitrary values to each death or serious injury and would predict the number of occurrences which would involve either death or serious injury. Finally, Ford would determine the cost to litigate such deaths and injuries. Syson testified that if the cost to repair the defect exceeded the other costs, Ford would not correct the defect.[67]

As described here, the procedure undertaken by Ford has all the earmarks of a standard economic analysis. In particular, the decision came down to an explicit comparison of the benefits of risk reduction and costs.[68] The task from a cost-benefit standpoint is to assess the cost of the safety improvement and to compare these costs to the benefits.[69] The benefits equal the expected number of injuries or deaths, multiplied by a dollar valuation of the adverse health outcomes. This also is the approach taken in the Learned Hand negligence formula. Although there can be disagreement regarding the dollar values attached to the benefit levels, the overall steps in the analysis are not controversial and follow standard economic blueprints for risk analysis. The valuation amounts are not, however, ideal, and I will examine their influence in the empirical study.

These examples illustrate my hypothesis that as a practical matter, once the accident has taken place, the jury may compare the small costs

---

[62] 922 S.W.2d at 579.
[63] *Id.*
[64] *Id.*
[65] *Id.* at 588–89.
[66] *Id.*
[67] *Id.*
[68] For an overview of how economists conduct cost-benefit analyses, see EDITH STOKEY & RICHARD ZECKHAUSER, A PRIMER FOR POLICY ANALYSIS 134–58 (1978).
[69] *Id.*; *see also* VISCUSI, FATAL TRADEOFFS, *supra* note 4, at 4–6.

associated with fixing an automobile with the catastrophic injury to the victim and find that they are not commensurable. The difficulty is that jurors are often unable to conceive of the pre-accident situation, where the tradeoff is not between the small cost of a design change to one vehicle and an identified life, but rather between costs of the design change across an entire model and a statistical expectation of possible fatalities that is based on engineering models rather than concrete evidence. The failure of jurors to view the accident from the situation of the company beforehand is a reflection of the more general phenomenon of hindsight bias. In effect, jurors may use the information conveyed by an accident in judging behavior as if they knew all along that a particular activity was risky, even though this had not been evident before the accident occurred.[70]

The issue that I will examine below using experimental evidence is whether risk analyses can be presented to jurors in a way that will make them better understand the importance of risk-cost tradeoffs. In doing so, I will examine not only different kinds of analyses the company might undertake but also different ways in which these analyses might be positioned within the context of litigation. Do jurors respond to the economic soundness of the analysis? Do they respond to the contextual setting of the analysis?

III. Experimental Results on Corporate Risk Analysis

To better understand how jurors respond to risk analysis evidence, I undertook a series of experiments involving jury-eligible citizens. Subjects for these experiments were recruited by a survey research firm. The test site used was Phoenix, Arizona for the first five scenarios considered.[71] I also examine additional, new results based on a sample from Austin, Texas.[72] In each instance, the samples were broadly representative of the adult population and were similar in terms of attributes such as educational background. Because the scenarios presented were identical except for the controlled manipulations, one would not expect different results because of the difference in locales.

The study participants were divided into different groups, each of which considered one case scenario involving an automobile accident. The scenarios differed in terms of whether the company performed an

---

[70] *See, e.g.*, Reid Hastie et al., *Juror Judgments in Civil Cases: Hindsight Effects on Judgments of Liability for Punitive Damages*, 23 Law & Hum. Behav. 597 (1999).

[71] The results for Scenarios 1 through 5 are discussed in Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 552–559. Additional statistical analysis appears in Viscusi, *Jurors*, *supra* note 6, at 115–27.

[72] The reason two survey waves were run was that the initial survey results indicated widespread resistance to corporate risk analysis. The sequel was intended to explore whether such biases could be reduced.

analysis, the character of the analysis, and other analysis-related aspects of the case. By making comparisons across different subject groups that considered different scenarios, it was possible to see whether juror acceptance of risk-cost studies could be increased by changing the character of the analysis or pertinent facts of the case. Other aspects of the scenarios were held constant in order to isolate the effects of these influences.

The basic case scenario, Scenario 1, involved an automobile company that manufactured cars. The defective vehicles resulted in a number of burn deaths:[73]

> A major auto company with annual profits of $7 billion made a line of cars with a defective electrical system design. This failure has led to a series of fires in these vehicles that caused 4 burn deaths per year. Changing the design to prevent these deaths would cost $16 million for the 40,000 vehicles affected per year. This safety design change would raise the price of cars $400 each. The company thought that there might be some risk from the current design, but did not believe that it would be significant. The company notes that even with these injuries, the vehicle had one of the best safety records in its class.
>
> The courts have awarded each of the victims' families $800,000 in damages to compensate them for the income loss and pain and suffering that resulted. After these lawsuits, the company altered future designs to eliminate the problem.

Thus in Scenario 1, the company did not conduct a risk-cost analysis. In Scenario 2, the company also failed to perform an analysis, but the cost per life saved was only $1 million. This information can be found in Table 1, which summarizes the seven different scenarios. In Scenario 2, the company could have prevented the deaths at a cost of $1 million each. The other scenarios had the higher cost per life saved of $4 million. At this higher cost, it becomes more difficult for the company to prevent the burn deaths. As a consequence, juries that are trying to attain an efficient result should be less likely to find fault with the company's safety decision.

Unlike Scenarios 1 and 2, Scenarios 3 through 7 all involved the company undertaking a risk analysis that compared the benefits and costs

---

[73] Some scenarios were run with a different number of deaths, but responses were not sensitive to this manipulation. The total lives lost took on values of 4 and 10. Apparently, this range of the number of deaths was not sufficiently great to affect respondents' assessments given the other aspects of the scenario. There were no statistically significant differences in the responses to the scenarios when the number of deaths was 10 rather than 4.

of the safety improvement. The risk analysis for Scenario 3 used the value of compensatory awards as the reference point:

> The company did a detailed analysis of the risk and estimated that 4 people would die each year. However, the company estimated that the liability cost would only be $800,000 per death based on the median award all industries pay for product-related fatalities. The company's estimate of the total court awards for the design problem was $3.2 million per year. As a result, the company estimated that the $4 million annual cost of making the change exceeded the estimated value of the court awards. The company concluded that it was cheaper not to adopt the safer design. The company notes that even with these injuries, the vehicle had one of the best safety records in its class.

For Scenario 4, the basis for valuing expected lives lost was the value of a statistical life used by the government:

> The company did a detailed analysis of the risk and estimated that 4 people would die on average per year. However, the cost to eliminate the risk was $4 million per fatality prevented. To determine whether the safety improvement was worthwhile, the company used a value of $3 million per accidental death, which is the value used by the National Highway Traffic Safety Administration in setting auto safety standards. The company estimated that the annual safety benefits of the safer design would be $12 million (4 expected deaths at $3 million per death), while the costs would be $16 million. As a result, the company believed that other safety improvements might save more lives at less cost. The company notes that even with these injuries, the vehicle had one of the best safety records in its class.

In each instance, the company presented with these figures realized that the costs exceeded the benefits, and as a consequence, chose not to make the safety improvement.

The risk-cost analysis in Scenario 3 valued the benefits of saving lives in the same manner as the various analyses undertaken by Ford and General Motors.[74] It used the value of compensatory damages for fatalities, which was set at $800,000. This particular figure was selected as a measure of the present value of the decedent's lost earnings, less the decedent's consumption share, plus compensation for pain and suffering.[75]

---

[74] *See supra* notes 31–36, 44–51 and accompanying text.

[75] For example, if wage growth is at the rate of interest, then a person earning $40,000 per year for an additional twenty years would earn a total of $800,000. If pain and suffer-

The company found that the added safety costs of $4 million per life exceeded these expected court awards. As a consequence, the company chose not to make the safety investment.

Using the value of court awards as a proxy for the benefits from lives saved has superficial appeal, particularly if one's focus is solely on the expected consequences in the judicial system. In particular, if past judicial outcomes provide a reliable measure of other companies' prospects in court, then a company assessing the prospective legal costs of dangerous products could use this value as a cost measure. It should be kept in mind, however, that this index ignores whatever value consumers attach to avoiding risks even if they expect that a well-functioning legal system will compensate them fully for their losses. In addition, the value of the compensatory award based on income loss alone surely will not compensate for non-pecuniary losses due to death or serious disability, because people value their own lives by more than the value of their income.[76]

Furthermore, the compensatory amount does not reflect the full value that potential victims place on preventing risks to life.[77] The court awards reflect compensation for the family of the deceased that is intended to meet some of the income losses resulting from the death.[78] From the standpoint of proper design of automobiles, the appropriate question is not what the company must pay out to compensate the survivors, but rather how valuable it is to the prospective accident victims to reduce the risks of death to themselves.[79] In particular, what is their willingness to pay for reduced risk, where this translates into a value of a statistical life that is generally substantially greater than the present value of lost earnings? For example, if people are willing to pay $300 to avoid a risk of 1/10,000, the value of the statistical life is $3 million, even though compensatory damages to a particular victim likely would fall below this figure.

To capture the higher, willing-to-pay value, the risk analysis in Scenario 4 used a $3 million value of life rather than the $800,000 value in Scenario 3. The $3 million figure was comparable to the highest values

---

ing compensation equals the deduction for the decedent's consumption, then the total compensatory award would be $800,000.

[76] *See surpa* note 39 and accompanying text.

[77] The discussion below draws on VISCUSI, FATAL TRADEOFFS, *supra* note 4, at 17–23, and VISCUSI, RATIONAL RISK POLICY 45–68 (1998).

[78] *See* Viscusi, *Misuses*, *supra* note 37, at 116–17.

[79] This willingness-to-pay principle is a basic component of benefit-cost frameworks. *See* Stokey & Zeckhauser, *supra* note 68. Indeed, the United States Office of Management and Budget recommends this approach for regulations throughout the federal government. *See* U.S. OFFICE OF MANAGEMENT AND BUDGET, REGULATORY PROGRAM OF THE UNITED STATES GOVERNMENT, APR. 1, 1992–MAR. 31, 1993 634 (1992). The Bush administration indicated that the same guidelines that were pertinent from 1992 through 1993 (Exec. Order No. 12,866, 50 Fed. Reg. 51735 (Oct. 4, 1993)) would be implemented under the current administration. *See* Memorandum from John D. Graham, to the President's Management Council, Presidential Review of Agency Rulemaking by OIRA (Sept. 20, 2001).

used by the United States Department of Transportation.[80] Moreover, the scenario stated that in adopting this value and in carrying out the analysis, the company followed the same procedure used by the United States government when setting automobile safety regulations. In this instance, the benefit value for saving lives was $3 million per life, and the cost per expected life saved was $4 million, so that on an efficiency basis it still was not worthwhile to adopt the safety improvement.[81] Because the company concluded that the annual benefits of $30 million would be less than the annual costs of $40 million, it chose not to make the safety improvement.

The next series of scenarios focused on how the company used the analysis. Scenarios 5 through 7 involved situations in which the company conducted the risk analysis as in Scenario 4, and the costs and benefits were the same as in Scenario 4. In Scenario 5, the company made a miscalculation and failed to realize that the benefits of risk reduction exceeded the costs.

> The company estimated that the safety benefits of the safer design would be $6 million (2 expected deaths at $3 million per death), while the costs would be $8 million. As a result, the company did not adopt a safer design since it believed that other safety improvements might save more lives at less cost. If, however, the company had assessed the risk accurately, the benefits of the safer design would have been $12 million, which exceeds the costs of the design change. The company notes that even with these injuries, the vehicle had one of the best safety records in its class.

If the company had not made this error, it would have fixed the car design defect because it would have realized that this was the efficient course of action. In Scenario 5, the company erred in concluding that improvements in safety were not warranted, whereas in Scenario 4 the company correctly concluded that the changes were not warranted on an efficiency basis.

Scenario 6 focused on the degree to which the company actually used the analysis in designing the automobile. In this scenario, the analysis was undertaken by a staff engineer, but the company claimed that the analysis did not play any role in the design of the car.

---

[80] *See* Matthew D. Adler & Eric A. Posner, *Implementing Cost-Benefit Analysis When Preferences Are Distorted*, 29 J. L. STUDIES 1105 app. at 1146 (2000).

[81] The point of view I advocate is that prospective lives saved should be given the value of a statistical life derived from people's willingness to pay for risk reduction. Thus, it is the values determined by the beneficiaries of the government policy that should count.

> The study indicated that other safety improvements might save more lives at less cost. The company said it never used the study in the design of the vehicle. It was an analysis by a staff engineer that did not play any role in the design decision.

This could occur in a large enterprise if the analyst was working in a division not responsible for the automobile design, or if the analysis appeared in a memorandum by a low-level employee that simply never made it to the corporate decision-makers. The question posed by this scenario was whether the fact that the analysis took place at all would influence juror behavior, as opposed to the situation where the company actually incorporated the analysis in its decision-making process. It should be noted that even if jurors are told that the company did not use the analysis in designing the car, they may not find this claim credible. Jurors might hypothesize that the company is attempting to minimize the role of the analysis for fear that jurors might penalize the company if they knew that it had based its decision to forego the safety improvement on risk-cost tradeoff considerations.

In an actual case situation, one would expect the company to mount an aggressive defense explaining why it conducted the analysis.[82] Scenario 7 examined whether the company could deflect some of the criticism of risk analyses by emphasizing the constructive role that such studies play. In Scenario 7, the respondents were told that the company had undertaken other such risk analyses in the past. These analyses had led to many major safety improvements in cars, but this particular modification did not pass muster on a risk-cost basis.

> The company had undertaken a series of similar risk analyses for other safety measures. These studies led to improved structural reinforcements in the doors, stability controls, and other improvements. But in this instance the company concluded that the extra costs to consumers were too great in comparison to the safety benefits. The company chose instead to make other design changes that might save more lives at less cost. The company notes that even with these injuries, the vehicle had one of the best safety records in its class.

The issue to be explored in this scenario was whether detailed articulation of the rationale underlying corporate risk analysis could reduce juror resistance to accepting the approach's legitimacy.

---

[82] This hypothesis is based on the apparent importance of analysis in influencing jury decisions. A responsible defense attorney presumably would attempt to explain the rationale for the analysis and how it played a constructive role in corporate decisions rather than accept the existence of an analysis as indicating that the firm displayed a reckless disregard for life.

Table 2 presents the results of respondents' assessments of the different scenarios. Each respondent considered only one of the seven scenarios. The number of respondents ranged from 96 for Scenario 5 to 104 for Scenario 6, or roughly 100 subjects each. Overall, 695 jury-eligible adults participated in the experiments.

In each case, respondents considered the corporate risk analysis scenario as well as other questions pertaining to their risk beliefs and attitudes. Focusing on the detailed written scenario description is, of course, a more limited exposure to a case than would occur in a trial context. Because this aspect of the study was common across all scenarios, any effect should be consistent across the results and only the differing aspects of the case descriptions should influence juror decisions. Actual trial experiences would provide a less effective experimental structure because differences in attorney presentation or other "live" factors would make it impossible to isolate the effects of the variables being studied.

The first question facing respondents after reading the scenario was whether they would award punitive damages. Each case scenario indicated that compensatory damages already had been awarded to the plaintiff. For the first two scenarios, in which no analysis was undertaken, the percentage awarding punitive damages was 85% for Scenario 1 and 92% for Scenario 2, a difference that was not statistically significant.

The appropriate reference point for judging the subsequent scenarios will continue to be how they differed from Scenario 1 as described above. Scenarios 3 through 7 each paralleled Scenario 1 in that the cost per life saved was $4 million, while they differed in terms of whether a risk analysis was undertaken and if so, what it entailed. The results for Scenarios 3, 4, and 5 with respect to whether punitive damages were awarded ranged from 93% to 95%, levels similar to the figures for Scenarios 1 and 2. It is noteworthy that when, in Scenario 6, the company did not use the analysis as part of its corporate decision-making, the percent favoring punitive damages dropped to 89%. Although this 89% figure still exceeds the 85% who would have awarded punitive damages for Scenario 1, the difference is not statistically significant. Remember that Scenario 1 was identical to Scenario 6, except that in the former, the company did no analysis whatsoever. This suggests that jurors cannot differentiate between the situation where a company disregards its risk analysis, and the situation where the company does not conduct a risk analysis at all.

Perhaps the most interesting results are those for Scenario 7, which suggest that it may be possible to communicate the usefulness of risk analysis to jurors. In that scenario, the percent favoring punitive damages dropped to 76%, which was the lowest for any scenario and was 9% below the value for Scenario 1.

What is striking about all of these results is that they indicate a very high willingness to impose punitive damages. When an automobile had a

known defect and the company chose not to repair the car, jurors had a substantial willingness to impose punitive damages irrespective of whether the safety measure had been subjected to a rigorous economic analysis. In fact, undertaking an analysis justifying the decision tended to slightly increase the percentage of jurors awarding punitive damages. The exception was Scenario 7, in which a very strong effort was made to convey the fact that such risk analyses had led to substantial safety improvements in the past. Irrespective of the different variations in the cases, at least three-fourths of all respondents favored punitive damages. When companies feasibly could have reduced known risks for a fairly reasonable cost, the respondents generally viewed the failure to do so as reckless behavior.

Whereas there was not a great deal of difference in the frequency with which the jury-eligible citizens awarded punitive damages for the different scenarios, the levels of the awards exhibited more variation. The last two columns of Table 2 provide information on the geometric mean and the median award for the different scenarios.[83] The first two scenarios—in which no analysis was performed at all—had similar results with respect to the award level: roughly $3 million for the geometric mean and $1 million for the median award. Underlying modest differences in the cost of promoting safety seem to have little effect on the level of punitive damages awards. In contrast, undertaking a risk analysis using an $800,000 value of life as in Scenario 3 boosted the geometric mean award to $4 million and the median award to $3.5 million. Thus, the risk analysis led to higher levels of punitive damages awards.

What happened when the company conducted its analysis with the $3 million value of life used by the National Highway Traffic Safety Administration to evaluate the effectiveness of regulations? The results for Scenario 4 indicate that doing so did not dampen jurors' concerns with the analysis, but rather increased both the geometric mean award level and the median award level, with the median reaching a high value of $10 million. It therefore appears that rather than making risk analyses more acceptable to jurors, use of a high value of life that reflects greater concern for safety on the part of the company serves as an anchor that boosts jury awards to a higher level.[84] Often in punitive damages contexts, jury verdicts are influenced by the desire to send the defendant company a message.[85] In order to encourage companies to change their

---

[83] The median values are more representative and less distorted by outliers than the mean award levels. If, for example, a respondent were awarded $1 billion in punitive damages, that amount would greatly influence the average across all respondents and would distort a measure of how the typical juror would perform. The median is also the measure of how the critical individual in a majority rule context would value the damages, which also may be important as a prediction of jury behavior.

[84] For similar conclusions based on analysis of Scenarios 1 through 5, see Viscusi, *Corporate Risk Analysis*, *supra* note 6, at 558–59.

[85] *See supra* note 42 and accompanying text.

policies, respondents may have reasoned that they had to assess a higher punitive damages value than the amount warranted by the value of life that the corporation used in its own decision-making. Otherwise, they would not be sending the company a message that the risk-cost analysis had undervalued human life. The high anchor value also may make jurors more reluctant to see the company make a sensitive tradeoff between risk to life and costs.

The results for Scenario 5, in which the company overestimated the costs, were similar to those for Scenario 4. From an economic standpoint, the company was actually more remiss in Scenario 5 than in Scenario 4 because the safety improvements in Scenario 5 were efficient on an economic basis. Nevertheless, there seems to be no evidence that the jurors want to punish the company for these errors in the cost estimates. There was no apparent effect on the frequency of punitive damages awards that could be traced to this difference. Rather, the jurors awarded damages similar to those awarded when the company took the efficient course of action. The jurors also awarded damages with similar frequency.

The final two scenarios directed at ameliorating the jurors' concerns each appeared to be somewhat successful in diminishing the damages awards. If the company did not in fact use the analysis in its automobile design, the geometric mean award dropped to $2.5 million, and the median award was $3 million. These results were similar in character to the results for the no-analysis Scenario 1. The difference between the two was that in Scenario 6, the company actually conducted a risk analysis, although that analysis never entered the decision-making process. Interestingly, the median award in Scenario 6 is above the median award in Scenarios 1 and 2, so it appears that doing an analysis that is not used affects the distribution of punitive damages awards adversely when compared against not doing an analysis at all.

Scenario 7 represented the attempt to decrease the punitive sanctions by elaborating on the constructive role that other analyses had played in car design decisions. This led to somewhat lower values for the geometric mean award than any of the other scenarios, as well as a median award level of $1 million that equaled those for Scenarios 1 and 2. It therefore appears that undertaking a responsible corporate risk analysis consistent with government regulatory practices may have a modest effect both on the frequency of punitive damages awards and on the value of the geometric mean award, reducing both figures. Still, the frequency and level of the sanctions remained quite substantial even when efforts were made to explain the nature of the analysis to the jury. Perhaps a company that successfully justifies its risk analysis as a safety-enhancing exercise can reduce the likely punitive damage sanctions. The effect is, however, modest, as jurors still remain willing to award damages. Despite the value of risk-cost analyses for obtaining efficient safety out-

comes, even well-educated jurors fail to reward companies for such efforts.

IV. Rules for Setting Punitive Damages[86]

The results above suggest that jurors may be more likely to award punitive damages against companies that perform risk analyses. Given jurors' failure to use risk analysis evidence constructively, one might ask whether a much tighter structure imposed on punitive damages awards could change their behavior. In particular, what if jurors were given explicit mathematical formulas for setting punitive damages? Such guidelines have, in fact, already been proposed in a set of model jury instructions by Polinsky and Shavell.[87] In this Part of the Article, I report on results of a study in which I gave subjects a case scenario and asked them to apply the Polinsky and Shavell damages formulas.

The essence of the Polinsky-Shavell damages approach is that the total damages should equal the amount of harm divided by the probability that this harm will be detected.[88] If there is a fifty percent chance that the harm will be detected, the total penalty should be twice the value of the damages incurred so that the expected damages borne by the offending company will equal the level of the harm inflicted. If the harm will be detected with certainty, then the amount of total damages should equal the amount of harm, or the compensatory damages amount. Punitive damages should equal zero. With the Polinsky-Shavell instructions, the court does not give the jurors the probability of detection; rather, the jury must assess this parameter. My experimental design simplifies this task by giving jurors the probability of detection. The basic case scenario involved toxic waste disposal. In order to link this behavior with the Polinsky-Shavell formula, the case included an explicit indication of the probability that the violation would be detected.

In the typical case scenario, the study participant was told that the Toxic Chemical Research Institute had twelve steel drums of dangerous chemicals to dispose of before a major production run. The shift manager was worried about the accumulating chemicals and wanted to find an easy way to dispose of them. The probabilities of detection of illegal disposal are summarized in the middle column of Table 3. In the first scenario, the company disposed of the chemicals in a nearby stream behind the plant, realizing that there was a twenty-five percent chance that the EPA inspector was going to be visiting the plant. Consequently, there was a seventy-five percent chance that the EPA inspector would not visit the

---

[86] See Viscusi, *Challenge of Punitive Damages*, *supra* note 7 for further discussion of these experimental results.

[87] *See* Polinsky & Shavell, *supra* note 7.

[88] *Id.* at 889.

plant, and that the company would not be caught. The scenario indicated that the EPA inspector did identify the spill and that the company was fined $100,000 to cover water treatment costs. The question was whether there should be punitive damages and, if so, what amount should be levied. Under the Polinsky-Shavell approach, the total damages should equal the total harm divided by the probability of detection, or $100,000 divided by 25%, or four times the economic loss, for a total of $400,000.

Scenario 2 was identical to Scenario 1 except that the detection probability was 1%. Thus, there was a 99% chance of escaping any penalty. Under the formula, the total damages levied for Scenario 2 should have been 100 times the economic loss. It should be noted that in Scenario 2, the detection probability is fixed by EPA inspection practices and cannot be manipulated by the polluter.[89]

Scenario 3 modified Scenario 2 in terms of the character of the company's behavior. The company loaded the chemical drums onto unmarked trucks and dumped the chemicals in a rural stream at 3 a.m. The manager did this because he hoped that this late-night dumping would reduce the risk of getting caught. Whereas the company in Scenario 2 also deliberately dumped chemicals in a stream that it knew might be inspected by the EPA, the low 1% probability of detection was outside of the control of the polluter. In contrast, in Scenario 3 the low probability of detection arose because the company engaged in stealthy behavior with the intention of decreasing the probability of detection. One therefore might expect the jurors to award higher penalties to punish the polluter's stealth.

Scenario 4 maintained the stealthy disposal assumption from Scenario 3 but added an anchoring factor. The study participants were told by the plaintiff's attorney that in order to send the company an appropriate signal to be more responsible, there should be a penalty of $50 million, or about 20% of the company's net worth. The attorney further argued that the minimum penalty should be $25 million, and, as a consequence, the penalty range to be considered by the jury should be between $25 million and $50 million.[90]

---

[89] Because full enforcement is costly, agencies generally do not have the resources to monitor every potential offender on a continuing basis, and detection probabilities will be less than one.

[90] Scenario 4 quoted the plaintiff's attorney:

> Your job as jurors is to impose a penalty which will make this corporation, and others, conduct their business in a way which protects the defenseless citizens of Texas who have no other way of getting the company to be responsible. This is your job. A penalty against this company has to be one that they will notice. It would not destroy this company or even cause them long term financial harm to impose a penalty on them of $50 million, about 20% of their net worth, or about two and one-half times their annual profit. Certainly a minimum penalty should be one year's profit, about $25 million, so the range you may want to consider is between $25 million, about one year's profit, and $50 million. I don't think that anything less than $25 million would have much effect as far as deterring them

Scenario 5 provided similar anchors, though in this case the quoted damages amount came from a newspaper article indicating that in a similar case in California, there had been a jury award of $50 million in punitive damages. The newspaper article also told the jurors that the appeals court reduced this award to $25 million.[91]

As in the case of the corporate risk analysis experiment, subjects considered only one of these different scenarios. The total number of participants in this experiment was 353. Each of the participants was given the Polinsky-Shavell jury instructions, which ask people to calculate three dollar amounts.[92] The first is the deterrence amount, which is inversely related to the probability of detection.[93] To assist in calculating this figure, Polinsky and Shavell provide a table so that jurors can determine the damages without any mathematical computation.[94] All participants in the study received a copy of this table.

The second portion of the instructions tells jurors to calculate a punishment amount to penalize blameworthy employees at the firm.[95] Qualitative guidelines are provided for considering this question, but the instructions do not give the jurors a punishment formula.[96] Jurors are then asked what amount of punitive damages should be awarded from the standpoint of punishment. Thus, this formula couples one subjective element with other objective components. Moreover, the subjective component is firmly grounded in detailed criteria for setting the punishment value.

The third component of the damages calculation involves finding the final size of the punitive damages award.[97] The instructions indicate that "[p]unitive damages should be an amount between the amount that you found appropriate for the purpose of deterrence and the amount that you found appropriate for the purpose of punishment."[98] Thus, the final num-

---

and getting them to be more careful in their operations.

[91] Scenario 5 stated:

> Before being placed on the jury you read about a similar case that took place in California. A jury there fined the company $50 million in punitive damages. However, the company appealed claiming the award was excessive. The punitive damages amount was reduced to $25 million by the appeals court in California. The company claimed that this amount was still too high and that it would continue to fight the award in court.

[92] Polinsky & Shavell, *supra* note 7, at 960.

[93] *See id.* at 960–61.

[94] *See id.* at 962.

[95] *Id.* at 961.

[96] For example, the instructions direct jurors to base the punishment amount on whether the defendant can identify blameworthy employees and whether shareholders or customers will be impacted negatively by punitive damaged levied on the defendant. *Id.*

[97] *See id.* at 962.

[98] *Id.*

ber for the punitive damages should fall between the deterrence value, which is governed by an explicit mathematical formula, and the punishment value, which is governed by a less precise set of standards.

How well did the jurors adhere to the instructions? The final column of Table 3 indicates the percentage of respondents who calculated the deterrence value of punitive damages correctly in view of the Polinsky-Shavell formula. For Scenario 1, 20% of the respondents correctly applied the formula, while 80% were unwilling or unable to do so. In the second scenario, in which the probability of detection was lower and the total value of damages consequently was boosted substantially according to the formula, the mock jurors appeared to be less willing to impose the substantial sanctions. The percentage of jurors who assessed a correct deterrence value for this scenario dropped to 11%. In contrast, if the behavior of the company was described as stealthy, people were more willing to impose the high deterrence punitive damage amounts prescribed by the formula. The results for Scenario 3 indicate that 21% of the respondents assessed a correct deterrence value.

If the instructions are to be effective, in the real world people must follow these guidelines rather than extraneous information presented as part of the case. Scenarios 4 and 5 considered the role of anchoring through statements by the plaintiff's attorney and information in newspaper articles on similar litigation. What we find is that once given these convenient anchors to latch onto, jurors tended to disregard the Polinsky-Shavell instructions to an even greater extent than before. Indeed, for Scenario 4, in which the plaintiff's attorney suggested an anchoring figure, only 7% of all respondents calculated the deterrence value correctly. This low statistic is particularly noteworthy because the computation required only minimal mathematical skills to execute.[99] This suggests that factors other than lack of mathematical ability led jurors to abandon the formula.[100]

The levels of damages awarded by the participants in the study are summarized in Table 4. For the first three scenarios, the damages from the standpoint of deterrence averaged approximately $3 million to $4 million, while the punishment values ranged between approximately $1 million and $6 million. Final award levels for these scenarios were approximately between $3 million and $6 million.[101]

---

[99] The damages amount was $100,000 in the case, and with a probability of escaping liability of 99%, the appropriate multiplier for punitive damages from the Polinsky-Shavell table is 99, leading to an optimal deterrence value of $9.9 million. Despite the simplicity of this calculation, 93% of the subjects reported deterrence values other than this correct amount.

[100] Controlling for personal characteristics and differences in the scenarios, only Scenario 4 exhibits statistically significant differences in terms of the probability of answering the deterrence questions correctly. See Viscusi, *Challenge of Punitive Damges*, *supra* note 7, at 339.

[101] The fact that Scenario 1 has a higher punishment value than Scenarios 2 and 3 and a

Matters change substantially when we examine Scenario 4, in which there was a plaintiff anchoring effect. These anchors increased the deterrence value by roughly an order of magnitude from the previous scenarios. Indeed, Scenario 3 was identical to Scenario 4 except for the presence of the plaintiff's anchor information. Yet, for the deterrence value, the punishment value, and the final punitive award, the responses for Scenario 4 were approximately 10 times greater than for Scenario 3. These effects were almost as strong, but not to the same extent, for Scenario 5.

What these results indicate is that not only do people fail to apply the deterrence value formula from the Polinsky-Shavell instructions correctly, but when presented with extraneous information, they do not appear constrained by the discipline offered by the formula and instead rely on more convenient anchors that may be either more compelling or easier to execute. These conclusions do not bode well for the Polinsky-Shavell instructions' potential to alter juror performance.

The role of anchoring raises the broader issue of whether anchors should or should not be provided to jurors. Meaningful anchors, such as those tied to the Polinsky-Shavell formulas, would be more useful than arbitrary anchors that have the appearance of rigor, but in fact are completely unrelated to how damages values should be set.

A final pertinent measure of the extent to which the damages instructions will be effective is whether people follow the third part of the instructions. That task is fairly simple, as respondents only have to find a final punitive damages award between the deterrence value and the punishment value. If jurors are willing to be disciplined at all by instructions, then presumably they should be able to follow such simple directions. Notwithstanding the simplicity of this mathematical task, an average of only 76% of the respondents were able to come up with a final award amount that was between their deterrence amount and their punishment value. The high percentage of jurors who failed in their task suggests that they may have been straying deliberately from the formula, rather than failing in its application due to poor math skills.

V. CONCLUSION

These studies send a bleak message to those with the objective of bringing about more efficient damages awards, because they suggest that jurors punish careful corporate decision-making on safety issues. Instead of interpreting risk analyses as evidence that defendants tried to meet their duty of care, jurors view such studies negatively when awarding punitive damages. Undertaking a risk analysis may both increase juror

---

punishment value in excess of the deterrence value seems largely due to the influence of outliers rather than systematic factors.

willingness to impose damages and also increase the size of the damages award. Jurors are unwilling to recognize that corporate risk analyses have a legitimate role to play and should not subject defendants to additional sanctions.

Explicit mathematical formulas also fail to induce efficient damages awards. Despite the straightforward nature of the Polinsky-Shavell instructions, most jurors failed to implement them correctly. In fact, many of the respondents disregarded even the most fundamental aspects of the instructions, such as the guidance that the total award should lie between the deterrence and punishment values.

A common theme of these results is that anchoring effects are often operative. If companies use higher values of life in their risk analyses, jurors seek to impose even higher levels of damages so as to send a signal to the corporation regarding their valuation of safety. Thus, companies that use a higher value of life that reflects a greater concern with safety may suffer the perverse consequence of increasing rather than decreasing their liability.

Not even explicit instructions or mathematical formulas can eliminate anchoring effects. When the plaintiff's attorney or a newspaper account suggested dollar amounts, the respondents based their damages awards on these values even though they had been told to follow the punitive damages instructions. These results highlight the dangers of attorneys naming dollar values that may anchor damages assessments on arbitrary amounts.

These findings suggest that there is no simple remedy for changing juror performance with respect to the award of punitive damages. These disappointing results do not imply that no effective solution exists. Perhaps, for example, one could rewrite the jury instructions in such a way that people would in fact choose to apply them when making their determination of the damages levels. Whether different sets of instructions could be implemented properly and would in fact be used by jurors cannot be determined without testing their performance in an experimental setting. What we can conclude at this juncture is that the current formulation is ineffective. Neither the Polinsky-Shavell instructions nor evidence of risk analysis enables jurors to focus on the efficiency of damages awards.

TABLE 1
SUMMARY OF CORPORATE RISK ANALYSIS CASE SCENARIOS

Scenario	Risk-Cost Analysis	Risk Value of Life	Cost per Life Saved	Other Case Components
1	No	N/A	$4 million	-
2	No	N/A	$1 million	-
3	Yes	$800,000	$4 million	-
4	Yes	$3 million	$4 million	-
5	Yes	$3 million	$4 million	Actual benefits exceeded
6	Yes	$3 million	$4 million	Company did not use analysis
7	Yes	$3 million	$4 million	Company based past improvements on analysis

TABLE 2
CORPORATE RISK ANALYSIS CASE RESULTS

Scenario	Summary Description	Sample Size	Percent Favoring Punitive Damages	Geometric Mean Award ($ millions)	Median Award ($ millions)
1	No analysis, $4 million cost per life	97	85	3.0	1.0
2	No Analysis, $1 million cost per life	97	92	2.9	1.0
3	Analysis, $800,000 value of life	97	93	4.0	3.5
4	Analysis, $3 million value of life	102	93	5.3	10.0
5	Analysis, costs over-estimated	96	95	4.5	10.0
6	Analysis, did not use	104	89	2.5	3.0
7	Analysis, past improvements noted	102	76	2.1	1.0

TABLE 3
PUNITIVE DAMAGES SCENARIOS

Scenario	Description	Sample Size	Percent Correct Deterrence Value
1	.25 detection probability	70	20
2	.01 detection probability	72	11
3	.01 probability, stealthy	72	21
4	.01 probability, stealthy, plaintiff anchoring	70	7
5	.01 probability, stealthy, newspaper anchoring	69	14

TABLE 4
AVERAGE DAMAGES VALUES FOR SCENARIOS
($ MILLIONS)

Survey Version	Deterrence Value	Punishment Value	Final Punitive Award
1	2.9	5.6	5.7
2	3.8	1.4	3.5
3*	3.7	2.4	3.2
4	34.1	29.2	34.8
5	20.1	16.4	22.3

* This sample is trimmed of one outlier who awarded damages in the billions.

NOTE

STATE EMPLOYEES AND SOVEREIGN IMMUNITY: ALTERNATIVES AND STRATEGIES FOR ENFORCING FEDERAL EMPLOYMENT LAWS

BRENT W. LANDAU*

Recently, the Supreme Court has interpreted the doctrine of state sovereign immunity as barring suits for damages by state employees against state employers for violations of their federal employment rights. As a result, the ability of federal employment laws to protect state employees has been greatly diminished. In this Note, Mr. Landau presents an extensive overview of some of the key federal employment statutes and of the Supreme Court's recent state sovereign immunity jurisprudence. After considering numerous alternatives, he argues that conditioning the receipt of federal funds on state waivers of immunity provides the best means of restoring the full protections of federal employment laws to state employees.

I. INTRODUCTION

For nearly twenty years, it has been clear that Congress can require state governments to comply with federal employment laws.[1] Accordingly, the approximately five million state employees are protected by federal statutes such as the Fair Labor Standards Act ("FLSA"),[2] the Age Discrimination in Employment Act ("ADEA"),[3] the Americans with Disabilities Act ("ADA"),[4] and the Family and Medical Leave Act ("FMLA").[5]

From a practical standpoint, however, recent Supreme Court decisions upholding state sovereign immunity have considerably diminished the ability of state employees to enforce their rights under these laws.[6] As

---

* Law Clerk to the Hon. Bruce W. Kauffman, United States District Court for the Eastern District of Pennsylvania. B.A., State University of New York at Binghamton, 1998; J.D., Harvard Law School, 2001. The author wishes to acknowledge the assistance of Professor David Westfall.

[1] *See* Garcia v. San Antonio Metro. Transit Auth., 469 U.S. 528 (1984).

[2] 29 U.S.C. §§ 201–219 (1994 & Supp. IV 1998).

[3] 29 U.S.C. §§ 621–634 (1994 & Supp. IV 1998).

[4] 42 U.S.C. §§ 12101–12213 (1994 & Supp. IV 1998).

[5] 29 U.S.C. §§ 2601–2654 (1994 & Supp. IV 1998).

[6] *See* Seminole Tribe of Fla. v. Florida, 517 U.S. 44 (1996); Alden v. Maine, 527 U.S. 706 (1999); Kimel v. Florida Bd. of Regents, 528 U.S. 62 (2000); Board of Trs. of Univ. of Ala. v. Garrett, 531 U.S. 356 (2001).

a result, state employees are left in a situation where they may have federal employment rights but no effective remedies.[7]

This Note examines alternatives and strategies available for the enforcement of federal employment laws against state governments.[8] It concludes that currently available alternatives, including state employment laws, agency enforcement of federal employment laws, actions for injunctive relief, and suits against state officials, are inadequate to protect the rights of state employees. It then evaluates potential strategies, including state waivers of immunity, valid abrogation of immunity, and *qui tam* actions. Because valid abrogation of immunity and *qui tam* actions would require significantly more congressional action and might still be invalidated by the Supreme Court, this Note concludes that the best strategy for state employees involves conditioning federal funds on state waivers of immunity. This is because protecting the federal employment rights of state employees is politically palatable, making Congress more likely to enact such provisions, and because the constitutional standard governing Congress's conditional spending power is not difficult to meet.

II. THE PROBLEM

This Part discusses the problem that confronts state employees as a result of the Supreme Court's recent state sovereign immunity jurisprudence. First, it introduces the state workers affected. It then gives an overview of the federal employment laws at issue. Finally, it summarizes the state of the law with regard to state employee suits for violations of federal employment laws.

---

[7] *Cf.* Marbury v. Madison, 5 U.S. (1 Cranch) 137, 163 (1803) ("The very essence of civil liberty certainly consists in the right of every individual to claim the protection of the laws, whenever he receives an injury . . . . The government of the United States has been emphatically termed a government of laws, and not of men. It will certainly cease to deserve this high appellation, if the laws furnish no remedy for the violation of a vested legal right."). Without effective remedies, the federal employment laws essentially become, for state employees, "only a promise to the ear to be broken to the hope, a teasing illusion like a munificent bequest in a pauper's will." Edwards v. California, 314 U.S. 160, 186 (1941) (Jackson, J., concurring).

[8] This Note does not extensively discuss the merits of the Supreme Court's recent state immunity jurisprudence. That has been done elsewhere, *see infra* note 187, and in any event is not necessarily the most pressing issue. For better or for worse, state immunity appears to be settled law, at least for the time being. Therefore, this Note focuses instead on practical options for state employees and maps out possible strategies for these employees to enforce their federal employment rights.

A. State Employees

State governments employ over 4.8 million people.[9] Of these, over 2.2 million, or about forty-six percent, work in education.[10] Another 600,000 people, or about twelve percent, work in hospitals or other health care jobs.[11] The rest are employed in a variety of fields, including corrections,[12] streets and highways,[13] police,[14] social insurance,[15] public welfare,[16] and others.[17]

In many ways, public employees resemble private employees. Indeed, the Public Employee Department of the AFL-CIO has referred to "the people who work in government service" as "the faces of America."[18] Public employers, just like private employers, "execute contracts with their employees whereby the employees agree to accept direction and render specified labor services toward achievement of employer-set goals in exchange for wages and salaries."[19] Wages in the public and private sectors are comparable.[20]

Despite these basic similarities, however, there are significant differences. Among these differences is the idea of "public goods." With public goods, "the end consumer cannot necessarily provide the service without participation by others and frequently cannot preclude others from its consumption even if able to provide it to himself or herself."[21]

---

[9] U.S. Bureau of the Census, *State Government Employment Data* (Mar. 1999), *available at* http://www.census.gov/govs/apes/99stus.txt (last visited Nov. 25, 2001). The exact number of state employees in March 1999, including both full-time and part-time workers, was 4,817,784. *Id.*

[10] *Id.* The March 1999 total of 2,229,340 educational workers includes employees who worked in higher education (2,064,788, including 669,363 instructional staff) and elementary and secondary education (55,033, including 38,632 instructional staff), as well as other educational workers. *Id.*

[11] *Id.* In March 1999, 435,344 state workers were employed in hospitals, and another 175,263 worked in other state health care jobs. *Id.*

[12] In March 1999, states employed 456,753 people in corrections. *Id.*

[13] In March 1999, states employed 251,019 people to work on streets and highways. *Id.*

[14] In March 1999, 99,686 people worked for states as police officers. *Id.*

[15] In March 1999, 95,593 people worked in social insurance administration. *Id.*

[16] In March 1999, 228,917 people worked in public welfare. *Id.*

[17] Other state jobs include central and financial administration (228,394 in March 1999), judicial and legal (148,463), transportation (36,399), parks and recreation (37,794), and natural resources and utilities (163,823). *Id.*

[18] PUBLIC EMPLOYEE DEPARTMENT, AFL-CIO, PUBLIC EMPLOYEES: FACTS AT A GLANCE ii (1995).

[19] DAVID A. DILTS ET AL., LABOR RELATIONS LAW IN STATE AND LOCAL GOVERNMENT 41 (1992).

[20] *See Debate Closed: Public and Private Pay is Comparable*, COLLECTIVE BARGAINING REPORTER (Fall 1996), *available at* http://www.afscme.org/wrkplace/cbr496_1.htm (last visited Nov. 25, 2001) [hereinafter *Debate Closed*].

[21] DILTS ET AL., *supra* note 19, at 7. Dilts's examples of "public goods" include national defense, highways, and schools. *See also* CLARENCE R. DEITSCH & DAVID A. DILTS, THE ARBITRATION OF RIGHTS DISPUTES IN THE PUBLIC SECTOR 2 (1990).

Only governments can provide public goods, and governments require employees to deliver these goods.[22]

In public employment, the "source of decision-making authority" is sovereignty, as opposed to property in the private sector.[23] Public employers are structured not to achieve efficiency but to maintain checks and balances, so authority is "diffused."[24] Also, public employers operate in a more "risk-free environment" than do private employers, because there is less of a need to earn profit or gain a market share.[25] Whatever the differences between public and private employees, state workers have legal disputes with their employers just like workers in the private sector.[26]

B. Federal Employment Laws

Four main federal employment laws applicable to state employees are affected by the Supreme Court's recent decisions on state immunity. They are the Fair Labor Standards Act (FLSA),[27] the Age Discrimination in Employment Act (ADEA),[28] the Americans with Disabilities Act (ADA),[29] and the Family and Medical Leave Act (FMLA).[30]

Other federal employment laws either are not affected by the Court's state immunity jurisprudence (these include Title VII[31] and the Equal Pay Act)[32] or are not applicable to state employees (such as the Occupational

---

[22] DILTS ET AL., *supra* note 19, at 8; *see also* DEITSCH & DILTS, *supra* note 21, at 2.

[23] DILTS ET AL., *supra* note 19, at 41–42.

[24] *Id.* at 42–43.

[25] *Id.* at 43–44. *Cf.* Scott Lehigh, *It's No Competition for Boston*, BOSTON GLOBE, Mar. 16, 2001, at A23 (arguing that inefficient government services should be contracted out to private companies). For more on the similarities and differences between public and private employers and employees, see HARRY T. EDWARDS ET AL., LABOR RELATIONS LAW IN THE PUBLIC SECTOR 45–52 (4th ed. 1991).

[26] It has been observed, however, that the litigation of these disputes in the public sector may differ significantly from private sector litigation. *See* PATRICIA A. BRANDIN & DAVID A. COPUS, IN DEFENSE OF THE PUBLIC EMPLOYER: CASE LAW AND LITIGATION STRATEGIES FOR DISCRIMINATION CLAIMS xxv–xxviii (1988). Specifically, public opinion and special interest groups, among other factors, may influence public employment litigation more than private employment litigation. *See id.*

[27] 29 U.S.C. §§ 201–219 (1994 & Supp. IV 1998).

[28] 29 U.S.C. §§ 621–634 (1994 & Supp. IV 1998).

[29] 42 U.S.C. §§ 12101–12213 (1994 & Supp. IV 1998).

[30] 29 U.S.C. §§ 2601–2654 (1994 & Supp. IV 1998).

[31] 42 U.S.C. §§ 2000e to 2000e-17 (1994 & Supp. IV 1998). Title VII prohibits employers from discriminating on the basis of race, color, religion, sex, or national origin. *See id.* § 2000e-2. Title VII applies to state governments. *See id.* §§ 2000e(b). See *infra* notes 181–184 and accompanying text for a discussion of why state employees' rights to sue their employers for Title VII violations are not affected by the state immunity doctrine.

[32] 29 U.S.C. § 206(d) (1994). Technically part of the Fair Labor Standards Act, the Equal Pay Act prohibits employers from wage discrimination on the basis of sex. *See id.* § 206(d)(1). The Equal Pay Act applies to state governments. *See id.* For a discussion of why state employees' rights to sue their employers for Equal Pay Act violations are not affected by the state immunity doctrine, see *infra* note 186 and accompanying text.

Safety and Health Act[33] and the Employee Polygraph Protection Act[34]). Were statutes such as the Occupational Safety and Health Act and the Employee Polygraph Protection Act amended to apply to state governments, they would likely face similar state immunity problems.[35]

1. Fair Labor Standards Act

The Fair Labor Standards Act ("FLSA")[36] regulates the compensation of workers. Among other requirements, the FLSA sets a minimum wage,[37] requires extra compensation for overtime work,[38] and prohibits retaliation against employees who assert their rights under the statute.[39]

When first enacted in 1938,[40] the FLSA applied only to private employers.[41] In 1966, Congress extended the reach of the FLSA to public employees in schools, hospitals, nursing homes, mental institutions, mass transit systems, and other institutions.[42] The FLSA was extended again by the Education Amendments of 1972 to cover employees in public preschools.[43]

In 1974, Congress again amended the FLSA, this time extending the statute's coverage to include almost all employees of state governments.[44] In 1976, the Supreme Court limited the effect of these amendments in *National League of Cities v. Usery*,[45] but the Court overruled itself nine

---

[33] 29 U.S.C. §§ 651–678 (1994 & Supp. IV 1998). The Occupational Safety and Health Act regulates working conditions and requires safe workplaces. *See, e.g., id.* § 654. By its terms, however, it does not apply to state governments. *See id.* § 652(5).

[34] 29 U.S.C. §§ 2001–2009 (1994 & Supp. IV 1998). The Employee Polygraph Protection Act prohibits lie detector use by many employers, *see id.* § 2002, but by its terms does not apply to state governments. *See id.* § 2006(a).

[35] *Cf. infra* Part II.C.

[36] 29 U.S.C. §§ 201–219 (1994 & Supp. IV 1998).

[37] *See id.* § 206(a). The current minimum wage is $5.15 per hour. *Id.* The FLSA allows payment of a subminimum wage, currently $4.25 per hour, to workers under twenty years of age for the first ninety days of work. *See id.* § 206(g).

[38] *See id.* § 207(a). The FLSA requires that employees working over forty hours a week be compensated at one and one-half times their regular rate for the overtime work. *Id.*

[39] *See id.* § 215(a)(4). The FLSA also regulates child labor, *see id.* § 212, homework, *see id.* § 211, and hot goods, *see id.* § 215(a)(1).

[40] Law of June 25, 1938, ch. 676, 52 Stat. 1060 (1938).

[41] *See* THE FAIR LABOR STANDARDS ACT 640 (Ellen C. Kearns ed., 1999) [hereinafter KEARNS].

[42] Fair Labor Standards Amendments of 1966, Pub. L. No. 89-601, 80 Stat. 830 (1966); *see also* KEARNS, *supra* note 41, at 23, 640. This extension was upheld as a valid exercise of the commerce power by the United States Supreme Court in *Maryland v. Wirtz*, 392 U.S. 183, 194 (1968).

[43] Pub. L. No. 92-318, 86 Stat. 235 (1972); *see also* KEARNS, *supra* note 41, at 25, 640.

[44] *See* Fair Labor Standards Amendments of 1974, Pub. L. No. 93-259, 88 Stat. 55 (1974); *see also* Kearns, *supra* note 41, at 28, 640.

[45] 426 U.S. 833 (1976) (holding that the amendments to the FLSA are unconstitutional insofar as they operate to displace the states' ability to structure employer-employee relationships in areas of traditional governmental functions).

years later in *Garcia v. San Antonio Metropolitan Transit Authority*.[46] Thus, as of 1985, the FLSA applied to nearly all employees of state governments.[47]

The FLSA does not, however, apply to state employers in exactly the same way as it applies to private employers. This is particularly so with regard to overtime requirements. In 1985, Congress amended the FLSA to allow state governments and their employees to agree to substitute compensatory time for monetary payments.[48] Also, the FLSA provides a partial exemption from its overtime requirements for employees of law enforcement and fire protection agencies.[49]

The FLSA minimum wage and overtime provisions do not apply to executive and professional employees who are paid salaries in either the public or private sectors.[50] This exemption is significant for public employees because of the high proportion of "white collar" workers in government employment.[51]

The FLSA provides for enforcement by individual employees and by the Department of Labor ("DOL"). Employees can bring civil actions for back wages and overtime,[52] as well as an additional equal amount of liq-

---

[46] 469 U.S. 528 (1985) (holding that state transit authorities are not immune from the minimum wage and overtime requirements of the FLSA).

[47] *See* KEARNS, *supra* note 41, at 33, 640–41. Still excluded from the FLSA were and are elected officials and their personal staffs, advisors, and policy-making appointees. *See* 29 U.S.C. § 203(e)(2)(C) (1994 & Supp. IV 1998); *see also* KEARNS, *supra* note 41, at 645–50.

[48] *See* Fair Labor Standards Amendments of 1985, Pub. L. No. 99-150, 99 Stat. 787 (1985) (codified at 29 U.S.C. § 207(o) (1994 & Supp. IV 1998)); *see also* KEARNS, *supra* note 41, at 659–60; WILLBORN ET AL., EMPLOYMENT LAW: CASES AND MATERIALS 662 (2d ed. 1998). Compensatory time is similar to paid leave and must be cashed out if not used by the end of the employment relationship. *See* KEARNS, *supra* note 41, at 660, 665. The agreement to receive compensatory time in lieu of overtime pay may be contained in a collective bargaining agreement or other agreement between the employer and employee. *See* 29 U.S.C. § 207(o)(2)(A) (1994 & Supp. IV 1998); *see also* KEARNS, *supra* note 41, at 661; WILLBORN ET AL., *supra*, at 662.

[49] *See* 29 U.S.C. § 207(k) (1994). This provision states that a public employer does not violate the overtime requirement so long as the employee works less than a specified number of hours in a 28-day period. *See id.* Under this provision, employees must work more hours before they receive overtime compensation. *See* KEARNS, *supra* note 41, at 687–88.

[50] *See* 29 U.S.C. § 213(a)(1) (1994) (exempting employees "employed in a bona fide executive, administrative, or professional capacity"); *see also* KEARNS, *supra* note 41, at 157–292, 673–82; WILLBORN ET AL., *supra* note 48, at 676. Overall, the FLSA covers eighty-eight percent of non-supervisory employees. *Id.* at 633.

[51] *See Debate Closed*, *supra* note 20 (noting that many public sector employees are "white collar" workers). By statute, the exemption applies specifically to teachers. *See* 29 U.S.C. § 213(a)(1) (1994); *see also* KEARNS, *supra* note 41, at 678.

[52] 29 U.S.C. § 216(b) (1994); *see also* KEARNS, *supra* note 41, at 1081–87; WILLBORN ET AL., *supra* note 48, at 676. Employees may bring collective actions on behalf of other similarly situated employees. 29 U.S.C. § 216(b) (1994); *see also* KEARNS, *supra* note 41, at 1081; WILLBORN ET AL., *supra* note 48, at 677. Collective actions are similar to class actions, but they are more limited because employees must "opt in" if they wish to participate. *See* KEARNS, *supra* note 41, at 21–22, 1159–69; WILLBORN ET AL., *supra* note 48, at 677.

uidated damages.[53] The DOL can bring civil actions on behalf of employees for back wages, overtime, and liquidated damages.[54] Additionally, the DOL can seek injunctions requiring the payment of back wages and overtime.[55] Finally, the FLSA provides for civil fines for repeated or willful violations[56] as well as criminal penalties.[57]

2. Age Discrimination in Employment Act

The Age Discrimination in Employment Act ("ADEA")[58] prohibits employers from discriminating against employees who are over forty years old on the basis of age.[59] The ADEA also prohibits retaliation against workers who assert their rights under the act.[60] It specifies, however, that in some cases age can be a bona fide occupational qualification, and that in such cases it is not illegal for an employer to act on the basis of age.[61]

---

[53] 29 U.S.C. § 216(b) (1994); *see also* KEARNS, *supra* note 41, at 1087–90; WILLBORN ET AL., *supra* note 48, at 677. The Portal-to-Portal Act of 1947, Pub. L. No. 80-99, 61 Stat. 84 (1947) (codified at 29 U.S.C. §§ 251–262 (1994 & Supp. IV 1998)), gave courts discretion as to whether to award liquidated damages when an employer acts in good faith. 29 U.S.C. § 260 (1994); *see also* KEARNS, *supra* note 41, at 21, 1223–75; WILLBORN ET AL., *supra* note 48, at 677. Successful employee-plaintiffs also receive an award of attorneys' fees and costs. 29 U.S.C. § 216(b) (1994); *see also* KEARNS, *supra* note 41, at 1097–1122; WILLBORN ET AL., *supra* note 48, at 677.

[54] 29 U.S.C. § 216(c) (1994); *see also* KEARNS, *supra* note 41, at 1078–79; WILLBORN ET AL., *supra* note 44, at 677. In such suits, no attorneys' fees or costs are awarded. *See id.*

[55] 29 U.S.C. § 217 (1994); *see also* KEARNS, *supra* note 41, at 1038–53; WILLBORN ET AL., *supra* note 48, at 677. Liquidated damages are not available in such proceedings. *See* KEARNS, *supra* note 41, at 1038–39. When the DOL seeks a restitutionary injunction, the rights of individual employees to file their own suits are terminated. 29 U.S.C. § 216(b) (1994); *see also* KEARNS, *supra* note 41, at 1157. In injunction proceedings brought by the DOL, unlike in civil actions for back pay, overtime, and liquidated damages, the FLSA gives no right to a jury trial. *See id.* at 1079–80.

[56] 29 U.S.C. § 216(e) (1994 & Supp. IV 1998) (providing for a penalty of up to $1000 for each violation); *see also* KEARNS, *supra* note 41, at 1002–08; WILLBORN ET AL., *supra* note 48, at 677.

[57] 29 U.S.C. § 216(a) (1994); *see also* KEARNS, *supra* note 41, at 1129–1138; WILLBORN ET AL., *supra* note 48, at 677. This provision is rarely used. *See id.*

[58] 29 U.S.C. §§ 621–634 (1994 & Supp. IV 1998).

[59] *Id.* §§ 623(a), 631(a); *see also* HENRY H. PERRITT, JR., CIVIL RIGHTS IN THE WORKPLACE § 1.04, at 12 (3d ed. 2001); BRANDIN & COPUS, *supra* note 26, at 104–06. Since 1986, there has been no upper age limit to ADEA coverage. *See* Age Discrimination in Employment Amendments of 1986, Pub. L. No. 99-592, 100 Stat. 3342 (1986); *see also* JOSEPH E. KALET, AGE DISCRIMINATION IN EMPLOYMENT LAW 5 (2d ed. 1990); BRANDIN & COPUS, *supra* note 26, at 100.

[60] 29 U.S.C. § 623(d) (1994); *see also* KALET, *supra* note 59, at 8, 77–78.

[61] 29 U.S.C. § 623(f)(1) (1994); *see also* KALET, *supra* note 59, at 88–95; BRANDIN & COPUS, *supra* note 26, at 106–09. A bone fide occupational qualification is one that is "reasonably necessary" to the job; for example, if necessary to ensure public safety. *See generally* W. Air Lines, Inc. v. Criswell, 472 U.S. 400 (1985). The ADEA also provides defenses for employer actions based on "reasonable factors other than age" and good cause, and actions pursuant to a bona fide seniority system or employee benefit plan. 29 U.S.C. § 623(f) (1994); *see also* KALET, *supra* note 59, at 88, 95–102; BRANDIN & COPUS, *supra* note 26, at 109–12.

When enacted in 1967, the ADEA applied only to private employers.[62] In 1974, Congress amended the ADEA to extend its coverage to state governments.[63] The ADEA exempts elected officials as well as their personal staffs and policy-making appointees.[64] It also allows state firefighting and law enforcement agencies to discriminate on the basis of age "pursuant to a bona fide hiring or retirement plan that is not a subterfuge to evade the purposes of the Act,"[65] and it provides for the involuntary retirement of "bona fide executive[s]" in both the public and private sectors.[66]

Enforcement of ADEA rights is similar to FLSA enforcement.[67] The Act allows for civil actions by individual employees[68] or the Equal Employment Opportunity Commission ("EEOC").[69] According to the statute, damages for ADEA violations are calculated as unpaid minimum wages or overtime compensation under the FLSA, and an equal additional amount of liquidated damages may be awarded for "willful violations."[70] Equitable remedies, such as reinstatement, may also be provided in appropriate circumstances.[71]

3. Americans with Disabilities Act

The Americans with Disabilities Act ("ADA")[72] prohibits discrimination on the basis of disability in employment, government services, housing, and public accommodations. Title I of the ADA bars disability

---

[62] *See* KALET, *supra* note 59, at 3; BRANDIN & COPUS, *supra* note 26, at 100.

[63] Fair Labor Standards Amendments of 1974, Pub. L. No. 93-259, 88 Stat. 55 (1974) (codified at 29 U.S.C. § 630(b) (1994)); *see also* KALET, *supra* note 59, at 3; BRANDIN & COPUS, *supra* note 26, at 100. The Supreme Court upheld the extension to state governments in *EEOC v. Wyoming*, 460 U.S. 226 (1983). *See also* KALET, *supra* note 59, at 3, 23.

[64] *See* 29 U.S.C. § 630(f) (1994). The Supreme Court has interpreted this provision to exclude appointed state judges from the Act's coverage. *See* Gregory v. Ashcroft, 501 U.S. 452, 464–67 (1991).

[65] 29 U.S.C. § 623(j)(2) (Supp. IV 1998); *see also* KALET, *supra* note 59, at 30; BRANDIN & COPUS, *supra* note 26, at 118–20.

[66] 29 U.S.C. § 631(c) (1994); *see also* KALET, *supra* note 59, at 29–30; BRANDIN & COPUS, *supra* note 26, at 103.

[67] *See* 29 U.S.C. § 626(a)–(b) (1994); KALET, *supra* note 59, at 44–47, 105–06; BRANDIN & COPUS, *supra* note 26, at 115–16.

[68] 29 U.S.C. § 626(c) (1994); *see also* KALET, *supra* note 59, at 36–43; BRANDIN & COPUS, *supra* note 26, at 114–16. Employees in such cases have a right to a jury trial. *See* 29 U.S.C. § 626(c) (1994).

[69] 29 U.S.C. § 626(b) (1994); *see also* KALET, *supra* note 59, at 43–45; BRANDIN & COPUS, *supra* note 26, at 115. When first enacted, the ADEA gave government enforcement responsibility to the DOL. *See* KALET, *supra* note 59, at 3. In 1978, this responsibility was transferred to the EEOC. *See* Reorganization Act of 1977, Pub. L. No. 95-17, 91 Stat. 29 (1977); *see also* KALET, *supra* note 59, at 3.

[70] 29 U.S.C. § 626(b) (1994); *see also* KALET, *supra* note 59, at 105–30; BRANDIN & COPUS, *supra* note 26, at 116. Punitive damages are not allowed under the ADEA. *See* KALET, *supra* note 59, at 130; PERRITT, *supra* note 59, § 10.4[D], at 15.

[71] *See* KALET, *supra* note 59, at 141–52.

[72] 42 U.S.C. §§ 12101–12213 (1994 & Supp. IV 1998).

discrimination in employment.[73] It also requires employers to make reasonable accommodations necessary for qualified employees with disabilities to perform their jobs.[74] Further, Title I of the ADA prohibits medical examinations and pre-employment inquiries about an applicant's disabilities[75] and forbids retaliation and coercion by employers.[76]

Title I of the ADA uses the same enforcement provisions as Title VII of the Civil Rights Act of 1964:[77] that is, enforcement by individual employees through private actions[78] or by the government.[79] Successful employees can obtain compensatory and punitive damages,[80] as well as equitable relief.[81] Unlike the FLSA and the ADEA, Title I of the ADA has applied to state governments since the law was enacted.[82]

---

[73] *Id.* at § 12112(a) (prohibiting discrimination in employment on the basis of disability against "qualified" individuals); *see also* PERRITT, *supra* note 59, §§ 11.01, 11.02[A], at 671–75; UNITED STATES COMMISSION ON CIVIL RIGHTS, HELPING EMPLOYERS COMPLY WITH THE ADA: AN ASSESSMENT OF HOW THE UNITED STATES EQUAL EMPLOYMENT OPPORTUNITY COMMISSION IS ENFORCING TITLE I OF THE AMERICANS WITH DISABILITIES ACT 24–27 (1998) [hereinafter HELPING EMPLOYERS]. The ADA defines "disability" as "a physical or mental impairment that substantially limits one or more . . . major life activities," including a history or perception of such an impairment. 42 U.S.C. § 12102(2) (1994). The statute does not protect applicants who are currently using illegal drugs. *See id.* § 12114(a).

[74] 42 U.S.C. § 12112(b)(5) (1994); *see also* PERRITT, *supra* note 59, §§ 11.02[A], 11.05, 11.06, at 672–73, 680–86; HELPING EMPLOYERS, *supra* note 73, at 25. Under the ADA, reasonable accommodations include "making existing facilities used by employees readily accessible to and usable by individuals with disabilities," 42 U.S.C. § 12111(9)(A) (1994), and "job restructuring, part-time or modified work schedules, reassignment to a vacant position, acquisition or modification of equipment or devices, appropriate adjustment or modifications of examinations, training materials or policies, the provision of qualified readers or interpreters, and other similar accommodations for individuals with disabilities," *id.* § 12111(9)(B). Employers are not required to take action that causes them "undue hardship." *Id.* § 12112(b)(5)(A).

[75] 42 U.S.C. § 12112(d) (1994); *see also* HELPING EMPLOYERS, *supra* note 73, at 25. Although the primary focus of this Note is current state employees, applicants for public employment are equally affected by the Supreme Court's state immunity cases.

[76] 42 U.S.C. § 12203 (1994); *see also* PERRITT, *supra* note 59, § 11.09, at 701–02.

[77] *See* 42 U.S.C. § 12117 (1994); *see also* PERRITT, *supra* note 59, § 11.10, at 702.

[78] 42 U.S.C. § 2000e-5(f)(1) (1994). Employees in such cases have a right to a jury trial. *Id.* § 1981a(c).

[79] 42 U.S.C. § 2000e-5(f)(1) (1994). For charges filed against private employers, the government agency empowered to bring suits is the EEOC; for claims against state governments, it is the Department of Justice. *See id. See also* HELPING EMPLOYERS, *supra* note 73, at 27.

[80] 42 U.S.C. § 1981a(a)(2) (1994); *see also* PERRITT, *supra* note 59, § 1.05[D], at 18. When the suit involves reasonable accommodation, the employer can avoid damage awards by establishing its good faith in providing an accommodation. 42 U.S.C. § 1981a(a)(3) (1994); *see also* PERRITT, *supra* note 59, § 1.05[D], at 18. Compensatory and punitive damages are not available for disparate impact claims. 42 U.S.C. § 1981a(a)(2) (1994); *see also* PERRITT, *supra* note 59, § 1.05[D], at 18.

[81] *See* 42 U.S.C. § 2000e-5(g)(1) (1994).

[82] *See id.* §§ 12111(5), 12202; *see also* PERRITT, *supra* note 59, § 11.02[C], at 675. In addition, unlike the ADEA, the ADA does not exclude elected officials and their staffs and appointees. *See* PERRITT, *supra* note 59, § 11.02[C], at 675.

Additionally, Title II of the ADA bars discrimination by public entities, including state governments.[83] Among other applications to state governments, Title II may prohibit disability discrimination in state employment. The Department of Justice ("DOJ") has so interpreted the ADA in its regulations,[84] but courts are split on the issue.[85]

The DOJ and the EEOC have issued a joint regulation describing their overlapping jurisdictions under the two titles.[86] This regulation states that the EEOC retains Title I charges against state governments for investigation, but that decisions on whether to litigate such charges are made by the DOJ Civil Rights Division.[87] When jurisdiction exists solely under Title II, the charge is immediately referred to the DOJ for investigation.[88]

4. Family and Medical Leave Act

The Family and Medical Leave Act ("FMLA")[89] provides for as many as twelve weeks of unpaid leave for eligible employees.[90] Employees are entitled to leave due to the birth of a son or daughter, placement of a foster child, or a serious health condition of the employee or of an

---

[83] 42 U.S.C. § 12131(1)(A) (1994); *see also* PERRITT, *supra* note 59, § 11.02[C], at 675–76. Essentially, Title II extends to all state governments the requirements of the Rehabilitation Act of 1973, 29 U.S.C. §§ 701–796 (1994 & Supp. IV 1998). *See* PERRITT, *supra* note 59, § 11.02[C], at 676. Section 504 of the Rehabilitation Act prohibits programs that receive federal funds from discriminating on the basis of disability. *See* 29 U.S.C. § 794 (1994 & Supp. IV 1998); BRANDIN & COPUS, *supra* note 26, at 122–24. The requirements and standards of the Rehabilitation Act are essentially interchangeable with those of the ADA. *See* PERRITT, *supra* note 59, § 11.03, at 679; WILLBORN ET AL., *supra* note 48, at 510. Indeed, Congress has mandated their consistent interpretation. *See* 29 U.S.C. § 794(d) (1994) ("The standards used to determine whether this section [of the Rehabilitation Act] has been violated . . . shall be the standards applied under . . . the Americans with Disabilities Act . . . ."); 42 U.S.C. § 12,117(b) (1994) ("The agencies with enforcement authority . . . shall develop procedures to ensure that administrative complaints filed under [the ADA] and under the Rehabilitation Act of 1973 are dealt with in a manner that avoids duplication of effort and prevents imposition of inconsistent or conflicting standards for the same requirements . . . .").

[84] *See* 28 C.F.R. § 35.140 (2000).

[85] *Compare* Bledsoe v. Palm Beach County Soil & Water Conservation Dist., 133 F.3d 816, 825 (11th Cir. 1998) (holding that Title II applies to discrimination in employment) *with* Zimmerman v. Oregon Dep't of Justice, 170 F.3d 1169, 1178 (9th Cir. 1999), *cert. denied*, 69 U.S.L.W. 3574 (U.S. Feb. 26, 2001) (holding that Title II does not apply to discrimination in employment). The Supreme Court has noted the split, but has not resolved it. *See* Board of Trs. of Univ. of Ala. v. Garrett, 531 U.S. 356 n.1 (2001). Given the DOJ's interpretation, *see Bledsoe*, 133 F.3d at 822, and Congressional intent, *see id.* at 821, the better view is that Title II applies to state employment discrimination.

[86] *See* 29 C.F.R. § 1640 (2000); *see also* HELPING EMPLOYERS, *supra* note 73, at 27.

[87] *Id.* § 1640.7(b).

[88] *Id.* § 1640.7(a)(2).

[89] 29 U.S.C. §§ 2601–2654 (1994 & Supp. IV 1998).

[90] *Id.* § 2612(a); *see also* KURT H. DECKER, FAMILY AND MEDICAL LEAVE: IN A NUTSHELL 23 (2000).

immediate family member.[91] During such a leave period, the employee's benefits must be maintained,[92] and at the conclusion of the leave, the employee must be restored to an equivalent position.[93] An employer may not interfere with the exercise of an employee's FMLA rights or discriminate against an employee for asserting such rights.[94] The FMLA applies to state governments.[95]

The FMLA is enforceable through civil actions by employees[96] and through civil or administrative actions by the DOL.[97] Recoverable damages include lost compensation or an amount equal to twelve weeks of compensation, plus interest, as well as an equal additional amount of liquidated damages and equitable relief.[98]

The FMLA does not modify the ADA,[99] but rather imposes distinct requirements.[100] Where their rights coincide, the statute providing greater protection to the employee is used.[101] The FMLA does not provide for an "undue hardship" defense like the ADA,[102] but it does limit the leave period to twelve weeks.[103] Thus, one of the statutes may provide more protection in a particular situation than the other.[104]

C. The Doctrine

Two strands of relevant doctrine affect state workers' employment rights. The first concerns the applicability of federal employment laws to the states, and the second concerns state immunity from suits filed by individuals to enforce these laws. As this section shows, although federal employment laws are applicable to state governments, individual employees are often unable to enforce their rights in court as a result of the Supreme Court's recent state immunity jurisprudence.

---

[91] 29 U.S.C. § 2612(a)(1) (1994); *see also* DECKER, *supra* note 90, at 37–40.

[92] 29 U.S.C. § 2614(c)(1) (1994); *see also* DECKER, *supra* note 90, at 47.

[93] 29 U.S.C. § 2614(a)(1) (1994); *see also* DECKER, *supra* note 90, at 48–51. Some highly compensated employees are exempted from this requirement. 29 U.S.C. § 2614(b) (1994); *see also* DECKER, *supra* note 90, at 49.

[94] 29 U.S.C. § 2615(a) (1994); *see also* DECKER, *supra* note 90, at 60.

[95] 29 U.S.C. § 2611(4)(A)(iii) (1994 & Supp. IV 1998); *see also* DECKER, *supra* note 90, at 31.

[96] 29 U.S.C. § 2617(a) (1994); *see also* DECKER, *supra* note 90, at 62–63.

[97] 29 U.S.C. § 2617(b) (1994); *see also* DECKER, *supra* note 90, at 63–67.

[98] *See* 29 U.S.C. § 2617(a)(1) (1994); *see also* DECKER, *supra* note 90, at 67–68.

[99] *See* 29 U.S.C. § 2651(a) (1994) ("Nothing in this Act . . . shall be construed to modify or affect any . . . law prohibiting discrimination on the basis of . . . disability.").

[100] *See* DECKER, *supra* note 90, at 189.

[101] *See id.*

[102] *See* 42 U.S.C. § 12,112(b)(5)(A) (1994); *see also supra* note 74.

[103] 29 U.S.C. § 2612(a) (1994).

[104] *See* WILLBORN ET AL., *supra* note 48, at 696.

1. Applicability of Federal Employment Laws to States

In *Garcia v. San Antonio Metropolitan Transit Authority*,[105] the Supreme Court held that Congress has the power, under the Commerce Clause,[106] to subject state and local governments to federal employment laws.[107] In reaching its decision, the Court overruled *National League of Cities v. Usery*,[108] a case decided nine years earlier in which the Court held that Congress could not impose FLSA requirements against state governments "in areas of traditional governmental functions."[109]

National League of Cities and its progeny forced the federal courts to distinguish between "traditional" government functions, which Congress could not regulate, and non-traditional functions, which it could regulate.[110] Surveying the distinctions that federal courts had made, the Court in *Garcia* stated that "[w]e find it difficult, if not impossible, to identify an organizing principle that places each of the cases in the first group on one side of a line and each of the cases in the second group on the other side."[111]

The Court held, therefore, that FLSA obligations could apply to state and local governments no matter what type of government function was involved.[112] Writing for the Court, Justice Blackmun[113] stated that "[w]e perceive nothing in the overtime and minimum-wage requirements of the FLSA, as applied to the San Antonio Metropolitan Transit Authority, that is destructive of state sovereignty or violative of any constitutional provision."[114] Indeed, the Court noted, "[t]he political process ensures that laws that unduly burden the States will not be promulgated."[115]

---

[105] 469 U.S. 528 (1985).

[106] U.S. CONST. art. I, § 8, cl. 3 ("The Congress shall have Power . . . . To regulate Commerce with foreign Nations, and among the several States, and with the Indian Tribes").

[107] *See* 469 U.S. at 557. *Garcia* dealt specifically with the application of the FLSA to a local government entity. The issues in *Garcia* applied to both state and local governments, although the two are treated differently with respect to sovereign immunity.

[108] 426 U.S. 833 (1976).

[109] *Id.* at 852. In *National League of Cities,* the Court held that the FLSA "impermissibly interfere[d] with the integral governmental functions" of state and local governments. *Id.* at 851. "If Congress may withdraw from the States the authority to make those fundamental employment decisions upon which their systems for performance of these functions must rest, we think there would be little left of the States' 'separate and independent existence.'" *Id.* (quoting Coyle v. Smith, 221 U.S. 559, 580 (1911) (internal quotation marks omitted)).

[110] *See Garcia*, 469 U.S. at 530, 538–39.

[111] *Id.* at 539.

[112] *See id.* at 556.

[113] Justice Blackmun was joined in the majority by Justices Brennan, White, Marshall, and Stevens.

[114] *Garcia*, 469 U.S. at 554.

[115] *Id.* at 556. Chief Justice Burger and Justices Powell, Rehnquist, and O'Connor dissented. Justice Powell protested that "today's decision effectively reduces the Tenth Amendment to meaningless rhetoric when Congress acts pursuant to the Commerce Clause." *Id.* at 560 (Powell, J., dissenting). Justice Rehnquist made the following predic-

2. Recent State Immunity Jurisprudence

In a recent line of cases, the Supreme Court seems to have reinterpreted the doctrine of state sovereign immunity so as to preclude state employees from suing their employers for damages under the FLSA, ADEA, ADA, and FMLA. These decisions explicitly eliminate the possibility of recovery under the ADEA and ADA while leaving a strong implication that monetary suits under the FLSA and FMLA may be barred as well. This section provides an overview of these cases.

a. Seminole Tribe

In 1996, the Supreme Court decided Seminole Tribe of Florida v. Florida.[116] In Seminole Tribe, the Court held that Congress did not have the power under Article I of the Constitution to abrogate a state's Eleventh Amendment immunity from suit.[117]

The Eleventh Amendment provides that "[t]he judicial power of the United States shall not be construed to extend to any suit in law or equity, commenced or prosecuted against one of the United States by Citizens of another State, or by Citizens or Subjects of any Foreign State."[118] Although the Eleventh Amendment's text refers only to diversity suits, the Supreme Court has long understood it to stand for a broad proposition of state immunity from suits for damages in federal courts.[119] In *Pennsylvania v. Union Gas*,[120] the Court had held that Congress, pursuant to its powers under Article I, could abrogate this immunity.[121]

---

tion: "I do not think it incumbent on those of us in dissent to spell out further the fine points of a principle that will, I am confident, in time again command the support of a majority of this Court." *Id.* at 580 (Rehnquist, J., dissenting). Although *Garcia* has never been explicitly overruled and remains valid law, the Court's recent state immunity jurisprudence indicates that, in large part and in practical terms, Justice Rehnquist's prediction has come true.

[116] 517 U.S. 44 (1996). The facts in *Seminole Tribe* concerned a provision of the Indian Gaming Regulatory Act, 25 U.S.C. § 2710(d)(7) (1994), that authorized suits against state governments. *See Seminole Tribe*, 517 U.S. at 47.

[117] *See Seminole Tribe*, 517 U.S. at 47. The Court also held that injunctive relief under *Ex parte Young*, 209 U.S. 123 (1908), was unavailable. *Id.* at 76. *Ex parte Young* actions are suits for declaratory or injunctive relief for continuing violations of federal law against state officials acting in their official capacities. *See* Alden v. Maine, 527 U.S. 706, 727 (1999).

[118] U.S. CONST. amend. XI.

[119] *See Seminole Tribe*, 517 U.S. at 54 (citing Hans v. Louisiana, 134 U.S. 1 (1890)). In *Hans*, the Court stated that to allow states to be sued by its own citizens for damages in federal court "was not contemplated by the Constitution when establishing the judicial power of the United States." 134 U.S. at 15.

[120] 491 U.S. 1 (1989).

[121] *See id.* at 19 (stating that Congress' Article I powers would be "incomplete without the authority to render States liable in damages").

In *Seminole Tribe*, the Court explicitly overruled *Union Gas*.[122] Writing for the Court, Chief Justice Rehnquist[123] stated that "the Eleventh Amendment prevents congressional authorization of suits by private parties against unconsenting states. The Eleventh Amendment restricts the judicial power under Article III, and Article I cannot be used to circumvent the constitutional limits placed on federal jurisdiction."[124]

Although *Seminole Tribe* held that Congress lacks power to abrogate state sovereign immunity under Article I, it recognized that Congress can do so pursuant to § 5 of the Fourteenth Amendment.[125] One year after *Seminole Tribe*, in *City of Boerne v. Flores*,[126] the Court defined the meaning of "appropriate" legislation under § 5.[127] In *City of Boerne*, the Court noted that § 5 confers "remedial" power and gives Congress only "the power 'to enforce,' not the power to determine what constitutes a constitutional violation."[128] The Court stated that for legislation to be "appropriate" under § 5, "[t]here must be a congruence and proportionality between the injury to be prevented or remedied and the means adopted to that end."[129]

b. Alden

Seminole Tribe affected only suits brought in federal court. Some statutes, including the FLSA, provide for suits by employees to enforce their statutory rights in both federal and state court.[130] The Supreme

---

[122] *Seminole Tribe*, 517 U.S. at 66.

[123] Chief Justice Rehnquist was joined in the majority by Justices O'Connor, Scalia, Kennedy, and Thomas.

[124] *Seminole Tribe*, 517 U.S. at 72–73. Although the Court in *Seminole Tribe* held that Congress could not use its power under Article I to abrogate state immunity in federal court, it did not disturb Congress' authority to abrogate immunity when acting pursuant to § 5 of the Fourteenth Amendment. The Fourteenth Amendment states, in part, "nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws," U.S. CONST. amend. XIV, § 1, and section 5 provides that "[t]he Congress shall have power to enforce, by appropriate legislation, the provisions of this article." U.S. CONST. amend. XIV, § 5. The Court's opinion in *Seminole Tribe* noted that "the Fourteenth Amendment, by expanding federal power at the expense of state autonomy, had fundamentally altered the balance of state and federal power struck by the Constitution." *Seminole Tribe*, 517 U.S. at 59 (citing Fitzpatrick v. Bitzer, 427 U.S. 445, 452–56 (1976)).

[125] *Seminole Tribe*, 517 U.S. at 59. Section 5 of the Fourteenth Amendment grants Congress the authority to enact legislation enforcing the provisions of the amendment. U.S. CONST. amend. XIV, § 5 ("The Congress shall have power to enforce, by appropriate legislation, the provisions of this article.").

[126] 521 U.S. 507 (1997). In *City of Boerne*, the Court held unconstitutional the Religious Freedom Restoration Act of 1993, 42 U.S.C. § 2000bb to 2000bb-4 (1994).

[127] *See City of Boerne*, 521 U.S. at 536.

[128] *Id.* at 519.

[129] *Id.* at 520. The congruence and proportionality test requires that there be a proportional relationship between the ends sought and the means employed to achieve those ends. *See id.* at 530.

[130] *See* 29 U.S.C. § 216(b) (1994) ("An action to recover the liability prescribed in ei-

Court rejected this enforcement option, however, in Alden v. Maine.[131] In Alden, the Court held that "the powers delegated to Congress under Article I of the United States Constitution do not include the power to subject nonconsenting States to private suits for damages in state courts."[132]

Justice Kennedy, writing for the Court,[133] stated that state immunity does not derive from the Eleventh Amendment.[134] Instead, "the States' immunity from suit is a fundamental aspect of the sovereignty which the States enjoyed before the ratification of the Constitution, and which they retain today . . . except as altered by the plan of the Convention or certain constitutional Amendments."[135] The Court stated that "[t]he Eleventh Amendment confirmed, rather than established, sovereign immunity as a constitutional principle; it follows that the scope of the States' immunity from suit is demarcated not by the text of the Amendment alone but by fundamental postulates implicit in the constitutional design."[136] Helping to confirm this principle, in the Court's view, was the Tenth Amendment, which provides that "[t]he powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people."[137]

Together, *Alden* and *Seminole Tribe* mean that individuals lack a judicial forum in which they may file damage suits against states, at least when the cause of action was created by Congress pursuant to its Article I powers.[138] Justice Kennedy recognized this, but he noted that "[t]he States and their officers are bound by obligations imposed by the Constitution and by federal statutes that comport with the constitutional design."[139] He went on to state that "[w]e are unwilling to assume the States

---

ther of the preceding sentences may be maintained against any employer (including a public agency) in any Federal or State court of competent jurisdiction . . . .").

[131] 527 U.S. 706 (1999).

[132] *Id.* at 712.

[133] In *Alden*, as in *Seminole Tribe* and other recent state sovereign immunity cases, the majority consisted of Chief Justice Rehnquist and Justices O'Connor, Kennedy, Scalia, and Thomas, and the dissenters were Justices Stevens, Souter, Ginsberg, and Breyer.

[134] *Alden*, 527 U.S. at 713.

[135] *Id.*

[136] *Id.* at 728–29. Although state sovereign immunity and Eleventh Amendment immunity are "not identical," 2 CIVIL ACTIONS AGAINST STATE AND LOCAL GOVERNMENT: ITS DIVISIONS, AGENCIES, AND OFFICERS § 10.3, at 56 (Jon L. Craig ed., 1992) [hereinafter CRAIG], recent Supreme Court decisions such as *Alden* have, to some extent, essentially merged the two doctrines into one, and they are frequently conflated by courts and commentators.

[137] U.S. CONST. amend. X; *see Alden*, 527 U.S. at 713–14. Note, however, that the Court in *Garcia v. San Antonio Metropolitan Transit Authority*, 469 U.S. 528, 557 (1985), rejected the Tenth Amendment rationale of *National League of Cities v. Usery*, 426 U.S. 833, 852 (1976), that federal employment laws could not apply to state governments "in areas of traditional governmental functions."

[138] As noted above, Congress may abrogate state sovereign immunity when it properly acts pursuant to § 5 of the Fourteenth Amendment. *See supra* note 124.

[139] *Alden*, 527 U.S. at 755.

will refuse to honor the Constitution or obey the binding laws of the United States."[140]

c. Kimel

The Court again addressed federal employment law in Kimel v. Florida Board of Regents.[141] In Kimel, the Court held that individual state employees could not seek damages from their employers for violations of the ADEA.[142] Justice O'Connor, writing for the familiar five-Justice majority,[143] applied a two-part test to determine whether Congress validly had abrogated the states' sovereign immunity from ADEA suits.[144]

First, Congress must have made its intention to abrogate immunity "unmistakably clear in the language of the statue."[145] Addressing this question, Justice O'Connor wrote that "the plain language of [the ADEA] clearly demonstrates Congress' intent to subject the States to suit for money damages at the hands of individual employees."[146]

Second, the statute must validly have abrogated state sovereign immunity.[147] Although the Court recognized that the extension of the ADEA to the states was within Congress' Article I powers,[148] it emphasized that, under *Seminole Tribe*, Article I alone cannot enable the abrogation of state sovereign immunity.[149] The *Kimel* Court stated that "if the ADEA rests solely upon Congress' Article I commerce power, the private petitioners in today's cases cannot maintain their suits against their state employers."[150]

Determining whether the ADEA validly abrogated state sovereign immunity pursuant to § 5, the Court applied the "congruence and proportionality" test of *City of Boerne*.[151] In doing so, it discussed claims of unconstitutional age discrimination under the Equal Protection Clause.[152]

---

[140] *Id.* In evaluating potential solutions to the problem faced by public employees in light of state sovereign immunity, this Note does not consider the good faith of the States as a sufficient safeguard. State employees do bring claims, and it can be assumed that, at least some of the time, the claims are meritorious. The object of this Note is to determine the best way for state employees to enforce their rights when, for whatever reason, state governments "refuse to . . . obey the binding laws of the United States." *Id.*

[141] 528 U.S. 62 (2000).

[142] *Id.* at 92.

[143] *See supra* note 133.

[144] *See Kimel*, 528 U.S. at 73.

[145] *Id.* (quoting Atascadero State Hosp. v. Scanlon, 473 U.S. 234, 242 (1985)).

[146] *Id.* at 74.

[147] *See id.* at 78.

[148] *See id.* (citing EEOC v. Wyoming, 460 U.S. 226, 243 (1983)).

[149] *See Kimel*, 528 U.S. at 78–79.

[150] *Id.* at 79.

[151] *See id.* at 82–83; *see also supra* note 129 and accompanying text.

[152] *See Kimel*, 528 U.S. at 83. The Equal Protection Clause of the Fourteenth Amendment states that "No State shall . . . deny to any person within its jurisdiction the equal protection of the laws." U.S. CONST. amend. XIV, § 1.

The Court reaffirmed that "age is not a suspect classification under the Equal Protection Clause."[153] It further stated that "[a]ge classifications, unlike governmental conduct based on race or gender, cannot be characterized as 'so seldom relevant to the achievement of any legitimate state interest that laws grounded in such considerations are deemed to reflect prejudice and antipathy.'"[154] Additionally, the Court noted, "old age does not define a discrete and insular minority because all persons, if they live out their normal life spans, will experience it."[155] Thus, it concluded, states may discriminate on the basis of age without violating the Equal Protection Clause so long as the use of age as a factor is "rationally related to a legitimate state interest."[156]

Turning to the ADEA, the Court declared that "[t]he Act, through its broad restriction on the use of age as a discriminating factor, prohibits substantially more state employment decisions and practices than would likely be held unconstitutional under the applicable equal protection, rational basis standard."[157] Although it noted that "[d]ifficult and intractable problems often require powerful remedies, and we have never held that § 5 precludes Congress from enacting reasonably prophylactic legislation,"[158] the Court found that "Congress never identified any pattern of age discrimination by the States, much less any discrimination whatsoever that rose to the level of constitutional violation."[159] Thus, the Court held the ADEA's abrogation of state sovereign immunity invalid.[160]

d. Garrett

The Supreme Court returned to the area of state immunity in February 2001 with its decision in Board of Trustees of the University of Ala-

---

[153] *Kimel*, 528 U.S. at 83 (citing Gregory v. Ashcroft, 501 U.S. 452, 470 (1991); Vance v. Bradley, 440 U.S. 93, 97 (1979); Massachusetts Bd. of Retirement v. Murgia, 427 U.S. 307, 313–14 (1976)). Suspect classifications (e.g., race) require strict scrutiny and must be necessary to a compelling government interest. Quasi-suspect classifications (e.g., gender) require heightened scrutiny and must be substantially related to an important government interest. All other classifications need only a rational basis; that is, they must be rationally related to a legitimate state interest. *See generally* GERALD GUNTHER & KATHLEEN M. SULLIVAN, CONSTITUTIONAL LAW 662–840 (13th ed. 1997).

[154] *Id.* (quoting Cleburne v. Cleburne Living Ctr., Inc., 473 U.S. 432, 440 (1985)).

[155] *Id.* (citing *Murgia*, 427 U.S. at 313–14). The reference to "discrete and insular minority" is from *United States v. Carolene Prods. Co.*, 304 U.S. 144, 152 n.4 (1938). A discrete and insular minority is one that is "in need of extraordinary protection from the majoritarian political process." *Murgia*, 427 U.S. at 313 (internal quotations omitted).

[156] *Kimel*, 528 U.S. at 83.

[157] *Id.* at 86. The Court held that the ADEA's bona fide occupational qualification defense, 29 U.S.C. § 623(f)(1) (1994); *see also supra* note 61 and accompanying text, did not equate the ADEA standard with the rational basis test. *Kimel*, 528 U.S. at 86–87.

[158] *Kimel*, 528 U.S. at 88.

[159] *Id.* at 89.

[160] *Id.* at 91.

bama v. Garrett.[161] In that case, the Court held that state employees cannot sue their employers for violations of Title I of the ADA.[162]

As in *Kimel*, the Court analyzed whether the Equal Protection Clause could support the statute's abrogation of state sovereign immunity.[163] Relying on *Cleburne v. Cleburne Living Center, Inc.*,[164] Chief Justice Rehnquist, writing for the Court, reaffirmed that disability was not a "quasi-suspect" classification for purposes of the Fourteenth Amendment.[165] With regard to rational basis review, the Court stated that "[a]lthough . . . biases may often accompany irrational (and therefore unconstitutional) discrimination, their presence alone does not a constitutional violation make."[166]

The Court went on to conclude that Congress, in subjecting states to private ADA suits, did not have sufficient evidence of a pattern of unconstitutional discrimination.[167] Referring to examples from the legislative record, he wrote that "even if it were to be determined that each incident upon fuller examination showed unconstitutional action on the part of the State, these incidents taken together fall far short of even suggesting the pattern of unconstitutional discrimination on which § 5 legislation must be based."[168]

Ultimately, the Court concluded, "[e]ven if it were possible to squeeze out of these examples a pattern of unconstitutional discrimination by the States, the rights and remedies created by the ADA against the States would raise the same sort of concerns as to congruence and proportionality as were found in *City of Boerne*."[169] Specifically, it suggested that "the accommodation duty far exceeds what is constitutionally required in that it makes unlawful a range of alternate responses that would be reasonable but would fall short of imposing an 'undue burden' upon the employer."[170]

---

[161] 531 U.S. 356 (2001).

[162] *Id.* at 372–74. The *Garrett* Court declined to address Title II of the ADA, although it noted that the courts of appeals are divided on the issue of whether Title II covers employment claims. *Garrett*, 531 U.S. at 360 n.1. Soon after deciding *Garrett*, the Supreme Court denied certiorari in a number of cases that raised Title II issues. *See, e.g.*, California Dep't of Motor Vehicles v. Dare, 121 S. Ct. 1187 (2001); Walker v. Missouri Dep't of Corr., 121 S. Ct. 1188 (2001).

[163] *Garrett*, 531 U.S. at 365.

[164] 473 U.S. 432 (1985).

[165] *Garrett*, 531 U.S. at 366.

[166] *Id.* at 367.

[167] *Id.* at 370.

[168] *Id.* The Court refused to consider examples of discrimination by local governments, noting that "the Eleventh Amendment does not extend its immunity to units of local government." *Id.* at 369; *see also* 2 CRAIG, *supra* note 136, § 10.15, at 83–84; GERALD E. FRUG, LOCAL GOVERNMENT LAW 1 (2d ed. Supp. 2000) ("After *Alden*, cities, unlike states, remain subject to damage suits in both state and federal court for violations of the Fair Labor Standards Act.").

[169] *Garrett*, 531 U.S. at 372.

[170] *Id.*; *see also supra* note 74 and accompanying text.

e. Summary

This line of cases indicates that state employees may not sue their employers for damages for violations of the FLSA, ADEA, ADA, or FMLA. The Supreme Court has explicitly held that states have immunity from suits brought by employees under the ADEA[171] and the ADA.[172] Although the Court has never expressly ruled on whether states have such immunity from FLSA suits, the result of Seminole Tribe suggests that they do.[173] Federal courts of appeals that have addressed this issue have uniformly interpreted the doctrine to preclude employee FLSA suits against states,[174] and the Supreme Court implicitly approved this result in Alden.[175]

Whether states may be sued by employees for FMLA violations is slightly more complicated, but the weight of authority indicates that here, too, states are immune. In adopting the FMLA, Congress found that "due to the nature of roles of men and women in our society, the primary responsibility for family caretaking often falls on women, and such responsibility affects the working lives of women more than it affects the working lives of men."[176] Thus, it can be argued that "despite the gender-neutral application of the FMLA, its primary purpose is to address the needs of an increasing number of women in the workforce confronted with the escalating demands that an altered family structure has engendered."[177] Accordingly, because gender requires heightened scrutiny under the Equal Protection Clause,[178] the leave requirements of the FMLA may be "appropriate" legislation pursuant to section 5 of the Fourteenth Amendment.[179] Some circuit courts, however, have rejected this argu-

---

[171] *See* Kimel v. Florida Bd. of Regents, 528 U.S. 62, 67 (2000).

[172] *See Garrett*, 531 U.S. at 374. The Court has not addressed whether states may be sued by employees for violations of Title II of the ADA or the Rehabilitation Act, but it is likely that the *Garrett* analysis would prohibit employee damage suits brought under these statutes.

[173] *Cf.* Seminole Tribe of Fla. v. Florida, 517 U.S. 44, 47 (1996).

[174] *See* Mills v. Maine, 118 F.3d 37 (1st Cir. 1997); Close v. New York, 125 F.3d 31 (2d Cir. 1997); Balgowan v. New Jersey, 115 F.3d 214 (3d Cir. 1997); Abril v. Virginia, 145 F.3d 182 (4th Cir. 1998); Wilson-Jones v. Caviness, 99 F.3d 203 (6th Cir. 1997), *modified on other grounds*, 107 F.3d 358 (6th Cir. 1997); Raper v. Iowa, 115 F.3d 623 (8th Cir. 1997); Quillin v. Oregon, 127 F.3d 1136 (9th Cir. 1997); Aaron v. Kansas, 115 F.3d 813 (10th Cir. 1997).

[175] *See* Alden v. Maine, 527 U.S. 706, 712 (1999).

[176] 29 U.S.C. § 2601(a)(5) (1994).

[177] Brian Ray, Note, *'Out the Window'? Prospects for the EPA and FMLA after* Kimel v. Florida Board of Regents, 61 OHIO ST. L.J. 1755, 1784 (2000).

[178] *See* United States v. Virginia, 518 U.S. 515, 532–33 (1996) (holding that gender classifications must have "exceedingly persuasive" justifications).

[179] *See* Ray, *supra* note 177, at 1785–91. Presumably, this argument would justify only the family leave provisions of the FMLA. The medical leave provisions of the law would probably be analyzed along the lines of *Garrett*, because they raise questions similar to those raised by the ADA.

ment, holding that in enacting the FMLA, Congress did not validly abrogate state immunity.[180]

It is important to note the limits of the Court's immunity jurisprudence. For example, under *Fitzpatrick v. Bitzer*,[181] states are not immune from employee suits brought to enforce Title VII of the Civil Rights Act of 1964.[182] The Court cited *Fitzpatrick* with approval in *Seminole Tribe*,[183] and lower courts have subsequently continued to follow *Fitzpatrick*.[184]

The Equal Pay Act ("EPA") is a more difficult case, but here, too, the better view is that states are not immune from suits brought by employees to enforce its provisions. The EPA is designed to address sex discrimination, and it is "proportional to the unconstitutional conduct it is designed to prevent."[185] Indeed, every circuit court that has considered the issue has held that states are not immune from EPA claims brought by state employees.[186]

Scholarly reaction to the Supreme Court's recent sovereign immunity jurisprudence has been mixed, but it has tended to be negative.[187] At

---

[180] *See* Hale v. Mann, 219 F.3d 61 (2d Cir. 2000); Kazmier v. Widmann, 225 F.3d 519 (5th Cir. 2000); Sims v. Univ. of Cincinnati, 219 F.3d 559 (6th Cir. 2000).

[181] 427 U.S. 445 (1976).

[182] *See id.* at 456–57.

[183] *Seminole Tribe*, 517 U.S. at 59; *see also* RICHARD H. FALLON ET AL., HART AND WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 111 (4th ed. Supp. 2000) (noting that "the majority in *Seminole* does not question the continued vitality of *Fitzpatrick*").

[184] *See* Ussery v. Louisiana, 150 F.3d 431 (5th Cir. 1998); Crum v. Alabama (*In re* Employment Discrimination Litig.), 198 F.3d 1305 (11th Cir. 1999). Although some commentators argue that the disparate impact provisions of Title VII are open to an immunity challenge, *see, e.g.*, Joanne C. Brant, Seminole Tribe, Flores *and State Employees: Reflections on a New Relationship*, 2 EMPLOYEE RTS. & EMP. POL'Y J. 175, 198 (1998), a better interpretation of the state immunity cases continues to make states subject to Title VII disparate impact suits. Although a showing of "discriminatory purpose" is required to make out an Equal Protection violation, *see* Washington v. Davis, 426 U.S. 229, 239 (1976), allowing for disparate impact suits is reasonable "prophylactic legislation." Kimel v. Florida Bd. of Regents, 528 U.S. 62, 88 (2000). As the *Crum* court held, "[t]he means used by Congress in the disparate impact provisions of Title VII, so closely aligned to the constitutional equal protection analysis, are neither incongruent with the purpose of preventing intentional discrimination in public employment, nor disproportionate to the injury to be avoided." 198 F.3d at 1323–24.

[185] *See* Ray, *supra* note 177, at 1777.

[186] *See Ussery*, 150 F.3d at 431; Kovacevich v. Kent State Univ., 224 F.3d 806 (6th Cir. 2000); Varner v. Illinois State Univ., 226 F.3d 927 (7th Cir. 2000), *cert. denied*, 121 S. Ct. 2241 (2001); O'Sullivan v. Minnesota, 191 F.3d 965 (8th Cir. 1999); Hundertmark v. Florida Dep't of Transp., 205 F.3d 1272 (11th Cir. 2000). This Note considers enforcement alternatives and strategies for the FMLA but not the EPA, even though the Supreme Court has not expressly addressed either statute in the context of state immunity, for two reasons. First, as noted in the text, the doctrine indicates that the EPA contains a valid abrogation of state sovereign immunity, whereas the FMLA may not. Second, every court of appeals that has considered the EPA and FMLA has reached these same conclusions. Therefore, while state employees retain the right to sue for money damages for EPA violations in every circuit, they cannot do so for FMLA violations in at least three.

[187] A discussion of the theoretical merits of the doctrine is beyond the scope of this Note. Such work has, however, been done by a number of scholars. *See generally* FALLON ET AL., *supra* note 183, at 119, 151–52 & nn.3–4; Carlos Manuel Vázquez, *Sovereign Im-*

this point, however, the Court seems to have settled on its current path. Therefore, new attention should be focused on ways to work within the doctrine to provide effective remedies for state employees seeking to enforce their federal employment rights.

III. Existing Alternatives

Before discussing possible new enforcement strategies, consideration of existing alternatives for the enforcement of employment rights is necessary. These alternatives include state employment laws, federal enforcement, suits for injunctive relief, and actions against state officers. If these alternatives are as effective as private damage suits against state employers for violations of federal laws, then perhaps no new enforcement strategies are needed.

Analysis will reveal, however, that the existing alternatives are insufficient. State laws do not provide equivalent protections or remedies, federal enforcement is not a viable option because of limited resources, injunctive actions will be largely ineffective in making employees whole, and individual official suits may be unavailable or otherwise inadequate. Accordingly, some action must be taken to provide state employees with effective remedies for their employers' violations of federal laws.

A. State Employment Laws

In both *Kimel* and *Garrett*, the Court suggested that state employees still have remedies for age and disability discrimination through state employment laws.[188] Likewise, many states have enacted laws regulating fair labor standards and family and medical leave. The question, therefore, is whether these state laws provide the same protections and remedies as their federal counterparts.

Every state has some statutory provision prohibiting state agencies from discriminating against their employees on the basis of age and disability.[189] These statutes, however, are far from uniform with respect to

---

munity, Due Process, and the Alden *Trilogy*, 109 YALE L.J. 1927, 1927 n.3 (2000).

[188] *See* Kimel v. Florida Bd. of Regents, 528 U.S. 62, 91 (2000) (stating that "State employees are protected by state age discrimination statutes, and may recover money damages from their state employers, in almost every State of the Union" and citing the statutes of forty-eight states); Board of Trs. of Univ. of Ala. v. Garrett, 531 U.S. 356, 374 n.9 (2001) (stating that "state laws protecting the rights of persons with disabilities in employment and other aspects of life provide independent avenues of redress").

[189] *See* ALA. CODE § 21-7-8 (1997) (disability); ALA. CODE § 25-1-22 (2000) (age); ALASKA STAT. § 18.80.220(1) (Michie Supp. 1999) (age and disability); ARIZ. REV. STAT. § 41-1463 (1999) (age and disability); ARK. CODE ANN. § 21-3-203 (Michie 1996) (age); ARK. CODE ANN. § 16-123-107 (Michie Supp. 1999) (disability); CAL. GOV'T CODE § 12941 (West 1992) (age); CAL. GOV'T CODE § 12940(a) (West Supp. 2001) (disability); COLO. REV. STAT.§ 24-34-402(1)(a) (Supp. 2000) (age and disability); CONN. GEN. STAT. § 46a-60 (2001) (age and disability); DEL. CODE ANN. tit. 19, § 711(a) (Supp. 1998) (age);

either their coverage or their remedies. Some state laws are ambiguous as to whether the state is a covered employer at all.[190] The age discrimination law in one state applies only to some executive branch employees, excluding many.[191] In another state, employers "subject to" the federal ADEA are excluded.[192]

In many states, employees are not even permitted to sue their state employers for age or disability discrimination.[193] In some states that do

---

DEL. CODE ANN. tit. 19, § 724(a) (1995) (disability); FLA. STAT. ch. 760.10(1) (1997) (age and disability); GA. CODE ANN. § 34-1-2 (1998) (age); GA. CODE ANN. § 45-19-29(1) (Supp. 2000) (age); GA. CODE ANN. § 34-6A-4(a) (1998) (disability); HAW. REV. STAT. § 378-2(1) (Supp. 1999) (age and disability); IDAHO CODE § 67-5909 (Michie 1995) (age and disability); 775 ILL. COMP. STAT. 5/2-102(A) (1998) (age and disability); IND. CODE § 22-9-2-2 (1998) (age); IND. CODE § 22-9-5-19 (1998) (disability); IOWA CODE § 216.6(1)(a) (1999) (age and disability); KAN. STAT. ANN. § 44-1113(a) (2000) (age); KAN. STAT. ANN. § 44-1009(1) (2000) (disability); KY. REV. STAT. ANN. 344.040 (Michie 1997) (age and disability); LA. REV. STAT. ANN. § 23:312(A) (West 1998) (age); LA. REV. STAT. ANN. § 23:323 (West 1998) (disability); ME. REV. STAT. ANN. tit. 5, § 4572(1)(A) (West Supp. 2000) (age and disability); MD. ANN. CODE art. 49B, § 16(a)(1) (Supp. 2000) (age and disability); MASS. GEN. LAWS ch. 151B, § 4(1C) (2000) (age); MASS. GEN. LAWS ch. 151B, § 4(16) (2000) (disability); MICH. COMP. LAWS § 37.2102 (1985 & Supp. 2000) (age); MICH. COMP. LAWS § 37.1202(1) (Supp. 2000) (disability); MINN. STAT. § 363.03(2) (2000) (age and disability); MISS. CODE ANN. § 25-9-149 (1999) (age and disability); MO. REV. STAT. § 213.055(1) (Supp. 2001) (age and disability); MONT. CODE ANN. § 49-2-303(a) (1999) (age and disability); NEB. REV. STAT. § 48-1004 (1998) (age); NEB. REV. STAT. § 48-1104 (1998) (disability); NEV. REV. STAT. § 613.330 (1999) (age and disability); N.H. REV. STAT. ANN. § 354-A:7 (Supp. 2000) (age and disability); N.J. STAT. ANN. § 10:5-12 (West Supp. 2000) (age); N.J. STAT. ANN. § 10:5-4.1 (West Supp. 2000) (disability); N.M. STAT. ANN. § 28-1-7(A) (Michie 2000) (age and disability); N.Y. EXEC. LAW § 296(1) (McKinney 2001) (age and disability); N.C. GEN. STAT. §§ 126-36, 143-422.2 (1999) (age); N.C. GEN. STAT. § 168A-5(a) (1999) (disability); N.D. CENT. CODE § 14-02.4-03 (1997) (age and disability); OHIO REV. CODE ANN. § 4112.02 (West Supp. 2001) (age and disability); OKLA. STAT. tit. 25, § 1302(A) (1987) (age and disability); OR. REV. STAT. § 659.030 (1999) (age); OR. REV. STAT. § 659.436 (1999) (disability); 43 PA. CONS. STAT. § 955(a) (Supp. 2000) (age and disability); R.I. GEN. LAWS § 28-5-7 (Supp. 1999) (age and disability); S.C. CODE ANN. § 1-13-80(A) (Law. Co-op. Supp. 2000) (age and disability); S.D. CODIFIED LAWS § 3-6A-15 (Michie 1994) (age); S.D. CODIFIED LAWS § 20-13-10 (Michie 1995) (disability); TENN. CODE ANN. § 4-21-401 (1998) (age); TENN. CODE ANN. § 8-50-103 (1993) (disability); TEX. LAB. CODE ANN. § 21.051 (Vernon 1996) (age and disability); UTAH CODE ANN. § 34A-5-016(1)(a) (Supp. 2000) (age and disability); VT. STAT. ANN. tit. 21, § 495(a)(1) (Supp. 2000) (age and disability); VA. CODE ANN. § 2.1-716 (Michie 1998) (age); VA. CODE ANN. § 51.5-41 (Michie Supp. 2000) (disability); WASH. REV. CODE ANN. § 49.44.090 (West Supp. 2001) (age); WASH. REV. CODE ANN. § 49.60.030 (West Supp. 2001) (disability); W. VA. CODE § 5-11-9 (1999) (age and disability); WIS. STAT. § 111.322 (1997-1998) (age and disability); WYO. STAT. ANN. § 27-9-105 (Michie 1999) (age and disability).

[190] *See, e.g.*, GA. CODE ANN. § 34-1-2 (2000); IND. CODE § 22-9-5-10 (1998).

[191] *See* S.D. CODIFIED LAWS § 3-6A-13 (Michie 1994).

[192] *See* IND. CODE § 22-9-2-1 (1998). This provision may create a "Catch 22" for Indiana state employees, *cf.* JOSEPH HELLER, CATCH-22 47 (1961), in that state employers are "subject to" the ADEA, and thus perhaps excluded from the state age discrimination statute, yet also immune from ADEA suits under *Kimel*.

[193] Private lawsuits are not permitted for disability discrimination in Alabama, *see* Averyt v. Doyle, 456 So. 2d 1096, 1098–99 (Ala. Civ. App. 1984), *rev'd on other grounds sub nom. Ex parte* Averyt, 487 So. 2d 912 (Ala. 1986); for age and disability discrimination in Delaware, see DEL. CODE ANN. tit. 19, § 712 (1995); for age discrimination in Georgia, see GA. CODE ANN. § 45-19-36 (1990); Calhoun v. Federal Nat'l Mfg. Ass'n, 823 F.2d

allow private lawsuits, jury trials are not permitted.[194] Even the substantive protections of the statutes are inconsistent across state lines. In one state, only workers with physical disabilities, not mental disabilities, are protected.[195] More generally, because each state statute is interpreted differently from the others, there is wide variation among the types of claims sustained, and the remedies awarded, in each state.[196]

Most states have their own minimum wage and overtime laws.[197] Some states, however, lack one or both of these.[198] A few states exempt

---

451, 455 (11th Cir. 1987); for age and disability discrimination in Illinois, see Armstrong v. Freeman United Coal Mining Co., 446 N.E.2d 296, 298 (Ill. App. Ct. 1983); for age and disability discrimination in Maryland, see Dillon v. Great Atl. & Pac. Tea Co., 403 A.2d 406, 409 (Md. Ct. Spec. App. 1979); for age and disability discrimination in Mississippi, see MISS. CODE ANN. § 25-9-149; for age discrimination in North Carolina, see N.C. GEN. STAT. § 126-36; Mullis v. Mechanics & Farmers Bank, 994 F. Supp. 680, 687 (M.D.N.C. 1997); for age discrimination in South Dakota, see S.D. CODIFIED LAWS § 3-6A-15; for age and disability discrimination in Utah, see UTAH CODE ANN. § 34A-5-107(15) (Supp. 2000); for age discrimination (other than discharge) in Virginia, see VA. CODE ANN. § 2.1-725 (Michie Supp. 2000); for age and disability discrimination in Wisconsin, see Bachand v. Conn. Gen. Life Ins. Co., 305 N.W.2d 149, 152–53 (Wis. Ct. App. 1981); and for age and disability discrimination in Wyoming, see WYO. STAT. ANN. § 27-9-107 (Michie 1999). Many of these states do allow for some sort of administrative remedy, but the procedures employed vary widely from state to state.

[194] *See, e.g.*, Smith v. Milliken & Co., 377 S.E.2d 916, 917–18 (Ga. Ct. App. 1989) (disability discrimination in Georgia); Smith v. ADM Feed Corp., 456 N.W.2d 378, 380–81 (Iowa 1990) (age and disability discrimination in Iowa); *cf.* 29 U.S.C. § 626(c) (1994) (allowing jury trials for ADEA suits); 42 U.S.C. § 1981a(c) (1994) (allowing jury trials for ADA suits).

[195] *See* ARIZ. REV. STAT. § 41-1461(4) (1999).

[196] A complete analysis of the many differences among the state statutes, and their variations from the federal laws, would not be practicable here. In a sense, however, this difficulty illustrates the central problem: there are a very large number of differences, both in the texts of the statutes and in their applications and interpretations. Indeed, entire annotations have been written collecting the case and statutory law of the states on various individual issues. *See, e.g.*, Cara Yates, Annotation, *Application of State Law to Age Discrimination in Employment*, 51 A.L.R.5th 1 (1997 & Supp. 2000); John E. Theuman, Annotation, *Right to Jury Trial in Action Under State Civil Rights Law*, 12 A.L.R.5th 508 (1993 & Supp. 2000); Jane M. Draper, Annotation, *Discrimination "Because of Handicap" or "On the Basis of Handicap" Under State Statutes Prohibiting Job Discrimination on Account of Handicap*, 81 A.L.R.4th 144 (1990 & Supp. 2000); Jane Massey Draper, Annotation, *Damages and Other Relief Under State Legislation Forbidding Job Discrimination on Account of Handicap*, 78 A.L.R.4th 435 (1990 & Supp. 2000); Jane Massey Draper, Annotation, *Handicap as Job Disqualification Under State Legislation Forbidding Job Discrimination on Account of Handicap*, 78 A.L.R.4th 265 (1990 & Supp. 2000); Jane M. Draper, Annotation, *Accommodation Requirement Under State Legislation Forbidding Job Discrimination on Account of Handicap*, 76 A.L.R.4th 310 (1990 & Supp. 2000); Brian H. Redmond, Annotation, *Award of Front Pay Under State Job Discrimination Statutes*, 74 A.L.R.4th 746 (1989 & Supp. 2000).

[197] *See* ALASKA STAT. § 23.10.065 (Michie 2000) (minimum wage); ALASKA STAT. § 23.10.060 (Michie 2000) (overtime); ARIZ. REV. STAT. § 23-391 (1995) (overtime); ARK. CODE ANN. § 11-4-210 (Michie Supp. 2001) (minimum wage); ARK. CODE ANN. § 11-4-211 (Supp. 1996) (overtime); CAL. LAB. CODE § 1182 (West 1989) (minimum wage); CAL. LAB. CODE § 204.3 (West Supp. 2000) (overtime); COLO. REV. STAT. § 8-6-106 (1994) (minimum wage and overtime); CONN. GEN. STAT. § 31-58(j) (2001) (minimum wage); CONN. GEN. STAT. § 31-76c (2001) (overtime); DEL. CODE ANN. tit. 19, § 902 (Supp. 2000) (minimum wage); GA. CODE ANN. § 34-4-3 (1998) (minimum wage); HAW. REV.

their own government employers from the minimum wage and overtime requirements.[199] Others exempt employees subject to the FLSA.[200] In some of the states that have their own minimum wage laws, that wage is

---

STAT. § 387-2 (1999) (minimum wage); HAW. REV. STAT. § 387-3 (1999) (overtime); IDAHO CODE § 44-1502 (Michie Supp. 2001) (minimum wage and overtime); 820 ILL. COMP. STAT. 105/4 (2000) (minimum wage); 820 ILL. COMP. STAT. 105/4a (2000) (overtime); IND. CODE § 22-2-2-4 (Supp. 2000) (minimum wage and overtime); IOWA CODE § 91D.1 (2001) (minimum wage); KAN. STAT. ANN. § 44-1203 (2000) (minimum wage); KAN. STAT. ANN. § 44-1204 (2000) (overtime); KY. REV. STAT. ANN. § 337.275 (Michie Supp. 2001) (minimum wage); KY. REV. STAT. ANN. § 337.285 (Michie Supp. 2001) (overtime); ME. REV. STAT. ANN. tit. 26, § 664 (West Supp. 2000) (minimum wage and overtime); MD. CODE ANN., LAB. & EMPL. § 3-413 (1999) (minimum wage); MD. CODE ANN., LAB. & EMPL. § 3-415 (1999) (overtime); MASS. GEN. LAWS ch. 151, § 1 (2000) (minimum wage); MASS. GEN. LAWS ch. 151, § 1A (2000) (overtime); MICH. COMP. LAWS § 408.384 (2001) (minimum wage); MICH. COMP. LAWS § 408.384a(8) (2001) (overtime); MINN. STAT. § 177.24 (2000) (minimum wage); MINN. STAT. § 177.25 (2000) (overtime); MO. REV. STAT. § 290.502 (2000) (minimum wage); MO. REV. STAT. § 290.505 (2000) (overtime); MONT. CODE ANN. § 39-3-404 (1999) (minimum wage) MONT. CODE ANN. § 39-2-405 (1999); NEB. REV. STAT. § 48-1203 (Supp. 2000) (minimum wage); NEV. REV. STAT. 608.250 (2000) (minimum wage); NEV. REV. STAT. 608.018 (2000) (overtime); N.H. REV. STAT. ANN. § 279:21 (1999) (minimum wage and overtime); N.J. STAT. ANN. § 34:11-56a4 (West 2000) (minimum wage); N.J. STAT. ANN. § 34:11-56a4.1 (West 2000) (overtime); N.M. STAT. ANN. § 50-4-22 (Michie Supp. 1999) (minimum wage and overtime); N.Y. LAB. LAW § 652 (McKinney Supp. 2001) (minimum wage); N.C. GEN. STAT. § 95-25.3(a) (1999) (minimum wage); N.C. GEN. STAT. § 95-25.4(a) (1999) (overtime); N.D. CENT. CODE § 34-06-03 (Supp. 2001) (minimum wage and overtime); OHIO REV. CODE ANN. § 4111.02(A) (West 2001) (minimum wage); OHIO REV. CODE ANN. § 4111.03 (West 2001) (overtime); OKLA. STAT. tit. 40, § 197.2 (1999) (minimum wage); OR. REV. STAT. § 653.025 (1999) (minimum wage); OR. REV. STAT. § 652.020 (1999) (overtime); 43 PA. CONS. STAT. § 333.104 (2001) (minimum wage and overtime); R.I. GEN. LAWS § 28-12-3 (2000) (minimum wage); R.I. GEN. LAWS § 28-12-4.1 (2000) (overtime); S.D. CODIFIED LAWS § 60-11-3 (Michie Supp. 2001) (minimum wage); TEX. LAB. CODE ANN. § 62.051 (Vernon 1996) (minimum wage); UTAH ADMIN. CODE 610-1-3(A) (2001) (minimum wage); UTAH CODE ANN. § 34-30-8 (1997) (overtime); VT. STAT. ANN. tit. 21, § 384 (1987 & Supp. 2000) (minimum wage and overtime); VA. CODE ANN. § 40.1-28.10 (Michie 1999) (minimum wage); WASH. REV. CODE ANN. § 49.46.020 (West Supp. 2001) (minimum wage); WASH. REV. CODE ANN. § 49.46.130 (West Supp. 2001) (overtime); W. VA. CODE § 21-5C-2(a) (Supp. 2000) (minimum wage); W. VA. CODE § 21-5C-3 (Supp. 2000) (overtime); WIS. STAT. § 104.04 (1997-1998) (minimum wage); WIS. STAT. § 103.01 (1997-1998) (overtime); WYO. STAT. ANN. § 27-4-202 (Michie 2001) (minimum wage); WYO. STAT. ANN. § 27-5-101 (Michie 2001) (overtime).

[198] States without either a minimum wage or overtime law are Alabama, Florida, Louisiana, Mississippi, South Carolina, and Tennessee. Arizona has an overtime law but no minimum wage law. States with a minimum wage law but without an overtime law are Delaware, Georgia, Iowa, Nebraska, New York, Oklahoma, South Dakota, and Texas.

[199] *See* IDAHO CODE § 44-1503 (Michie 2001) (exempting State from minimum wage and overtime requirements); VT. STAT. ANN. tit. 21, § 384(b)(7) (2000) (exempting State from overtime requirement); WYO. STAT. ANN. § 27-4-201(a)(iv)(D) (Michie 2001) (exempting State from minimum wage and overtime requirements).

[200] *See* ARK. CODE ANN. § 11-4-203 (Michie Supp. 2001); GA. CODE ANN. § 34-4-4(c) (1998); HAW. REV. STAT. § 387-1 (1999); IND. CODE § 22-2-2-3 (1998); KAN. STAT. ANN. § 44-1202(d) (2000); MO. REV. STAT. § 290.502; N.C. GEN. STAT. § 95-25.14(a)(1) (2000); TEX. LAB. CODE ANN. § 62.151 (Vernon 1996); VA. CODE ANN. § 40.1-28.9(B)(12) (Michie 1999). For a discussion of why such language might result in a problem for state employees, see *supra* note 192.

lower than the federal minimum.[201] Furthermore, many states do not allow recovery of liquidated damages.[202]

Thirty states have a statute or regulation providing some sort of family or medical leave to their employees.[203] There is, however, wide variation in the scope of the leaves provided by these states.[204] Twenty states have no laws or regulations providing for family or medical leave at all.[205]

While it is true that many states have their own laws prohibiting age and disability discrimination, regulating fair labor standards, and providing for family and medical leave, the above discussion suggests that these laws fall short of providing protection and remedies equivalent to those of the FLSA, ADEA, ADA, and FMLA. Coverage, substantive re-

---

[201] States with a minimum wage below $5.15 per hour, *cf.* 29 U.S.C. § 206(a) (1994), are Kansas, *see* KAN. STAT. ANN. § 44-1203 ($2.65), New Mexico, *see* N.M. STAT. ANN. § 50-4-22 ($4.25), Ohio, *see* OHIO REV. CODE ANN. § 4111.02(A) ($4.25), and Texas, *see* TEX. LAB. CODE ANN. § 62.051 ($3.35).

[202] *See, e.g.*, ARIZ. REV. STAT. § 23-391; ARK. CODE ANN., § 11-4-218 (Michie Supp. 2001); CAL. LAB. CODE § 98.7(d) (West Supp. 2001); COLO. REV. STAT. § 8-6-118 (1986); DEL. CODE ANN. tit. 19, § 911 (1995); KAN. STAT. ANN. § 44-1211 (2000); KY. REV. STAT. ANN. 337.550 (Michie 1995); MD. CODE ANN., LAB. & EMPL. § 3-427 (1999); NEV. REV. STAT. § 608.260 (2000); OHIO REV. CODE ANN. § 4111.10 (West 2001); OR. REV. STAT. § 653.055 (1999); 43 PA. CONS. STAT. § 333.113 (2001); R.I. GEN. LAWS § 28-12-19 (2000); VT. STAT. ANN. tit. 21, § 395 (2000); VA. CODE ANN. § 40.1-28.12 (Michie 1999); WYO. STAT. ANN. § 27-4-204 (Michie 2001); *cf.* 29 U.S.C. § 216(b) (1994) (allowing liquidated damages in FLSA actions).

[203] *See* ALA. ADMIN. CODE R. 670-X-14-.01 (1995); ALASKA STAT. § 23.10.500 (Michie 2000); ARIZ. ADMIN. CODE 2-5-411, 2-5-412 (2000); CAL. GOV'T CODE § 12945.2 (West Supp. 2001); 4 COLO. CODE REGS. § 801(P-5-24)–(P-5-38) (2001); CONN. GEN. STAT. § 31-51ll (2001); DEL. CODE ANN. tit. 29, § 5116 (1997); FLA. STAT. ch. 110.221 (Supp. 2001); HAW. REV. STAT. § 398-3 (1999); IDAHO ADMIN. CODE § 15.04.01.243 (2001); ILL. ADMIN. CODE tit. 80, § 420.645 (2000); IOWA ADMIN. CODE r. 581-14.4 (1999); KAN. ADMIN. REGS. 1-9-6 (2000); ME. REV. STAT. ANN. tit. 26, § 844 (West Supp. 2000); MD. CODE ANN., STATE PERS. & PENS., § 9-501 (1997); MONT. CODE ANN. § 2-18-606 (1999); NEV. REV. STAT. 284.360 (1999); N.J. STAT. ANN. § 34:11B-4 (West 2000); N.C. ADMIN. CODE tit. 25, r. 1E.0305 (Nov. 1995); N.D. CENT. CODE § 54-52.4-02 (1989); OKLA. STAT. tit. 74, § 840-2.22 (Supp. 2001); OR. REV. STAT. § 659.476 (1999); R.I. GEN. LAWS § 28-48-2 (1995); S.C. CODE ANN. §§ 8-11-40, -155 (Law. Co-op. Supp. 2000); S.D. ADMIN. R. 55:01:22:08.02 (2000); UTAH ADMIN. CODE 477-8-9 (2001); VT. STAT. ANN. tit. 21, § 472 (Supp. 2000); WASH. REV. CODE § 49.12.270 (1990 & Supp. 2001); W. VA. CODE § 21-5D-4 (1996 & Supp. 2000); WIS. STAT. § 103.10 (Supp. 2000).

[204] For example, a number of states do not allow for all the types of leave authorized by the FMLA. *See, e.g.*, DEL. CODE ANN. tit. 29, § 5116 (permitting adoption leave only); MONT. CODE ANN. § 2-18-606 (permitting adoption and paternity leave only); *cf.* 29 U.S.C. § 2612(a)(1) (1994) (permitting leave for the birth of a son or daughter, placement of a foster child, or a serious health condition of the employee or a family member). Other states provide for less leave than the FMLA. *See, e.g.*, HAW. REV. STAT. § 398-3 (allowing four weeks of leave); WIS. STAT. § 103.10 (allowing two weeks of leave for medical reasons or to care for a child, spouse, or parent, and up to eight weeks of leave for other reasons); *cf.* 29 U.S.C. § 2612(a)(1) (1994) (allowing up to twelve weeks of leave).

[205] The following states have no leave statutes or regulations comparable to the FMLA: Arkansas, Georgia, Indiana, Kentucky, Louisiana, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Nebraksa, New Hampshire, New Mexico, New York, Ohio, Pennsylvania, Tennessee, Texas, Virginia, and Wyoming.

quirements, and remedies vary from state to state, and, in many jurisdictions, state employees have no right to sue for violations.

Additionally, states can amend their own laws at any time.[206] Where, as here, the issue is the liability of the states themselves for illegal employment actions taken against their employees, some states may be motivated to amend their statutes to limit their exposure further.[207]

Finally, any argument that state employment laws accomplish all the purposes of the federal laws misses one important federal goal. The reason for having federal employment laws at all is to establish national standards and to ensure minimum employment rights. Congress has already made the policy judgment, in passing each of these statutes, that it is better to have federal laws that apply to employers nationwide than to rely on a patchwork of state laws.

B. Federal Enforcement

State sovereign immunity limits the ability of individual employees to sue their state employers, but it does not affect suits brought by the federal government.[208] Indeed, the FLSA, ADEA, ADA, and FMLA all provide for enforcement actions by the United States.[209] The Supreme Court, in its state immunity decisions, has several times emphasized the existence of federal enforcement remedies.[210]

As an initial matter, however, the federal government, by itself, simply could not handle the large volume of cases state employees bring against their employers. The United States currently litigates only a very small proportion of federal employment cases filed in federal district courts. Of 1,914 FLSA cases filed in 2000 (against both public and private defendants), the United States was plaintiff in only 382, and of

---

[206] Indeed, state legislatures can and do change the rules of litigation, even with regard to pending lawsuits. *See, e.g.*, Brent W. Landau, Recent Legislation, *State Bans on City Gun Lawsuits*, 37 HARV. J. ON LEGIS. 623 (2000). The fact that some of these state provisions are in the form of regulations rather than statutes provides even less assurance that they will not be altered.

[207] One way states might do this, in addition to changing the substantive law, is by requiring litigants to file suit in a claims court. *See* Brant, *supra* note 184, at 179. In a claims court, plaintiffs might be required to forfeit their right to a jury trial, to bifurcate their liability and damage claims, *see id.*, or to comply with certain pre-action notice provisions, *see* 1 CRAIG, *supra* note 136, § 5.1, at 332–33.

[208] *See* Alden v. Maine, 527 U.S. 706, 755 (1999) ("In ratifying the Constitution, the States consented to suits brought by other States or by the Federal Government.") (citing Principality of Monaco v. Mississippi, 292 U.S. 313, 328–29 (1934)).

[209] *See* 29 U.S.C. §§ 216(a), 216(c), 216(e), 217 (1994 & Supp. IV 1998) (FLSA); 29 U.S.C. § 626(b) (1994) (ADEA); 42 U.S.C. §§ 12117, 2000e-5(f)(1) (1994) (ADA); 29 U.S.C. § 2617(b) (1994) (FMLA).

[210] *See, e.g.*, Board of Trs. of Univ. of Ala. v. Garrett, 531 U.S. 356, 374 n.9 (2001); *Alden*, 527 U.S. at 759–60; Seminole Tribe of Fla. v. Florida, 517 U.S. 44, 71 n.14 (1996).

19,670 employment civil rights cases filed (including under the ADA and ADEA), the United States was plaintiff in only 425.[211]

Indeed, there seems to be almost universal agreement that the federal government, with its current resources, could not adequately fill the gap left by the absence of private enforcement.[212] If the federal government were to take a more active role in enforcement of the federal employment laws against state employers, a massive enlargement of its efforts would be necessary.[213] To this end, Justice Souter, dissenting in *Alden*, wrote that "unless Congress plans a significant expansion of the National Government's litigating forces to provide a lawyer whenever private litigation is barred by today's decision and *Seminole Tribe*, the allusion to enforcement of private rights by the National Government is probably not much more than whimsy."[214]

Such an expansion, given the current political climate, seems unlikely.[215] Issues of simple fiscal conservatism aside, the huge increase in federal spending that would be necessary to enforce federal employment laws effectively on behalf of state employees hardly seems to make sense. Professor Meltzer has observed that "Congress may reasonably doubt that federal government resources are wisely used to pursue litigation against state agencies when a private rightholder's interest is great but the public interest is small."[216] As Justice Souter argued, "it is implausible to claim that enforcement by a public authority without any incentive beyond its general enforcement power will ever afford the private right a traditionally adequate remedy."[217]

---

[211] *See* ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS, JUDICIAL BUSINESS OF THE UNITED STATES COURTS: 2000, Table C-2, *available at* http://www.uscourts.gov/judbus2000/appendices/c02sep00.pdf (last visited Mar. 16, 2001) (suits where the United States was the defendant omitted). Professor Meltzer has noted that such "statistics do not distinguish suits against states from those against other defendants. Since private plaintiffs sometimes file in state court but the United States virtually never does, these federal court statistics may slightly understate the ratio of private-plaintiff cases to U.S.-as-plaintiff cases." Daniel J. Meltzer, *State Sovereign Immunity: Five Authors in Search of a Theory*, 75 NOTRE DAME L. REV. 1011, 1022 n.50 (2000).

[212] *See* Roger C. Hartley, *The* Alden *Trilogy: Praise and Protest*, 23 HARV. J.L. & PUB. POL'Y 323, 373–74 & nn.254–57 (1999) (citing comments from the Congressional Record, Solicitor General, DOL, and the National Association of Police Organizations).

[213] *See* FALLON ET AL., *supra* note 183, at 149 (stating that the national executive could enforce federal obligations on states only "at considerable public expense").

[214] *Alden*, 527 U.S. at 810 (Souter, J., dissenting).

[215] *See, e.g.*, Anne E. Kornblut, *Bush to Demand Belt-Tightening Despite Surplus*, BOSTON GLOBE, Feb. 19, 2001, at A1 (reporting that President Bush will order most federal agencies, including the DOJ, to restrain spending).

[216] Meltzer, *supra* note 211, at 1022. Professor Meltzer suggests that "the fact that a violator of federal law is a state agency [does not] necessarily provide a reason for public enforcement." *Id.*

[217] *Alden*, 527 U.S. at 810 (Souter, J., dissenting). Additionally, one commentator has noted that there is more than a little irony present in the Rehnquist Court's emphasis on federal enforcement as a viable alternative:

In addition to the implausibility of a well-funded federal enforcement regime, there is evidence that private enforcement simply works better. It has been observed that "[g]overnment bureaucracies may respond far less quickly than private lawyers to shifting demand for enforcement action, and even decentralized government bureaucracies are likely to be less convenient and flexible than are private lawyers."[218]

For example, one commentator has described DOL enforcement of the FLSA as having "marked weaknesses."[219] Compliance officers can investigate only two percent of covered employers and discover only one-fifth of all underpayments.[220] Moreover, because of a lack of resources, there is a pressure to settle instead of litigate, and suits for liquidated damages, which require jury trials, are rarely brought.[221] This leads to the conclusion that "the enforcement procedures through the Department of Labor do not deter violators."[222] For this reason, private suits have advantages over DOL enforcement.[223]

With regard to the ADA, there is further reason to be concerned about government enforcement. As noted above, the DOJ, rather than the EEOC, brings ADA suits against state governments.[224] The United States Commission on Civil Rights, however, has reported as follows:

> [The] EEOC and the Department of Justice (DOJ) do not coordinate their ADA litigation activities very well. This is particularly true with regard to employment issues under title II (State and local employers). DOJ only litigates a small portion of its State and local employment cases. Part of the problem is that there is not much employment expertise at DOJ.[225]

> The image of a bulging and expensive federal bureaucracy of government attorneys descending on state agencies to enforce federal law is hardly the federalism of Justices Black or Powell or, one would have thought, Justices Rehnquist, Kennedy, or Scalia. Moreover, an effective federal enforcement effort that substantially duplicates the pre-*Seminole Tribe/Alden* private effort does little to alleviate the states' autonomy and fiscal apprehensions, which so manifestly inform the Court's reasoning in *Alden*.

Hartley, *supra* note 212, at 376.

[218] Meltzer, *supra* note 211, at 1023; *see also* Clyde Summers, *Effective Remedies for Employment Rights: Preliminary Guidelines and Proposals*, 141 U. PA. L. REV. 457, 480–81 (1992) (arguing that government enforcement is ineffective and slow).

[219] Summers, *supra* note 218, at 492.

[220] *Id.*

[221] *Id.* at 494.

[222] *Id.*

[223] *See id.* at 498. Even with private enforcement, estimates are that at least ten percent of employees clearly covered by the FLSA do not receive any premium pay for overtime. *See* WILLBORN ET AL., *supra* note 44, at 683.

[224] *See* 29 C.F.R. § 1640 (2000); HELPING EMPLOYERS, *supra* note 73, at 27.

[225] HELPING EMPLOYERS, *supra* note 73, at 206. This problem led the Commission to

Accordingly, although the state sovereign immunity doctrine creates a need for increased federal enforcement, such enforcement is not likely to provide equivalent remedies for aggrieved state employees. Due to limited federal resources, the government is an inadequate substitute for private plaintiffs in actions against state employers. Additionally, federal enforcement in general is likely to be less effective than private actions by individual employees and their attorneys.

C. Injunctive Relief

The doctrine of state sovereign immunity does not prohibit suits for declaratory and injunctive relief from continuing violations of federal law.[226] This remedy was originally authorized by the Supreme Court in *Ex parte Young*[227] in 1908, and was recently reaffirmed.[228] In such suits, state officers are the named defendants, but the states themselves are the real parties in interest.[229]

Since *Ex parte Young* actions cannot provide aggrieved employees with money damages, they are of limited usefulness as an effective remedy.[230] Also, some cases will not be appropriate for injunctions because either the employer's conduct is not ongoing or because the employer has since come into compliance with the law.[231] Nevertheless, suits for injunctive relief do provide individual state employees with some way to enforce federal employment laws.[232]

---

recommend that "EEOC and the Department of Justice should coordinate better by working together to ensure that charges relating to State and local employers are handled appropriately. The Department of Justice, for example, should make greater use of EEOC's employment expertise through better cooperative efforts and information-sharing programs." *Id.* at 265–66.

[226] *See* Alden v. Maine, 527 U.S. 706, 757 (1999).

[227] 209 U.S. 123, 166 (1908).

[228] *See* Idaho v. Coeur d'Alene Tribe of Idaho, 521 U.S. 261, 269 (1997) ("We do not, then, question the continuing validity of the *Ex parte Young* doctrine."). In *Coeur d'Alene*, an Indian tribe brought suit against various Idaho state officials for a declaratory judgment to establish its title to Lake Coeur d'Alene. *See id.* at 264–65.

[229] Hartley, *supra* note 212, at 378.

[230] *See* Brant, *supra* note 184, at 178 ("[R]elief under *Ex parte Young* cannot encompass backpay or compensatory damages, precisely those forms of relief that motivate many plaintiffs to endure the hardships of the litigation process."); Hartley, *supra* note 212, at 379 ("[W]hile this alternative may end ongoing violations of federal law, it does not grant compensation for past wrongs and does not deter noncompliance by the states until injunctive relief is granted."); Meltzer, *supra* note 211, at 1016–17 ("[S]tate officials who act as Holmesian bad men would have little to lose, and much to gain, by resisting compliance with the FLSA unless and until an injunction is obtained."); *see also* 2 CRAIG, *supra* note 136, § 12.21, at 95–96. Justice Holmes described the "bad man" as someone "who cares only for the material consequences which such knowledge [of the law] enables him to predict," as opposed to the "good [man], who finds his reasons for conduct, whether inside the law or outside of it, in the vaguer sanctions of conscience." Oliver W. Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 459 (1897).

[231] *See* Meltzer, *supra* note 211, at 1016.

[232] For more information on equitable relief in employment cases, see generally 2

There are two additional problems, however, with using *Ex parte Young* actions in lieu of the private suits now forbidden by the Supreme Court. First, some statutes do not permit individual employees to bring suits for injunctive relief. In particular, the FLSA[233] and FMLA[234] allow only the DOL to seek injunctions.[235]

Second, the Supreme Court recently may have limited the availability of *Ex parte Young* actions. In *Seminole Tribe*, the Court held that "where Congress has prescribed a detailed remedial scheme for the enforcement against a state of a statutorily created right, a court should hesitate before casting aside those limitations and permitting an action against a state officer based upon *Ex parte Young*."[236] It is possible, therefore, that courts will be reluctant to permit *Ex parte Young* actions where a statute provides for its own detailed remedial scheme, as the federal employment laws do.[237]

Thus, for several reasons, *Ex parte Young* actions do not provide an adequate way for state employees to enforce their federal employment rights. Most importantly, such suits cannot result in money damages for injured employees or address past conduct by an employer. Additionally, employees are prohibited from bringing actions for injunctions under the

---

CRAIG, *supra* note 136, § 14.15–.19, at 494–503.

[233] *See* 29 U.S.C. § 211(a) (1994) ("Except as provided in section 212 of this title [relating to child labor], the Administrator [Secretary of Labor] shall bring all actions under section 217 of this title [relating to injunction proceedings] to restrain violations of this chapter."); *see also* Meltzer, *supra* note 211, at 1016 & n.29 (arguing that it would be necessary to amend the FLSA to allow employees to seek injunctions).

[234] *See* 29 U.S.C. § 2616(a) (1994) (giving DOL the investigative authority provided by the FLSA); 29 U.S.C. § 2617(d) (1994) (providing for an "action [for an injunction] by the Secretary" of Labor).

[235] Employees are allowed to seek injunctions under the ADA, *see* 42 U.S.C. § 12117(a) (1994) (providing that Title VII enforcement procedures apply to the ADA); 42 U.S.C. § 2000e-5(g)(1) (1994) (allowing a court to issue injunctions and order affirmative relief in Title VII actions brought by employees), and the ADEA, *see, e.g.*, Criswell v. W. Airlines, Inc., 709 F.2d 544, 558 (9th Cir. 1983).

[236] Seminole Tribe of Fla. v. Florida, 517 U.S. 44, 74 (1996). One commentator has argued that the Court's conclusion on this point is "simply unsupported." Vicki Jackson, Seminole Tribe, *The Eleventh Amendment, and the Potential Evisceration of* Ex parte Young, 72 N.Y.U. L. REV. 495, 517 (1997). The Court noted that it did "not hold that Congress *cannot* authorize federal jurisdiction under *Ex parte Young* over a cause of action with a limited remedial scheme," only that "Congress did not intend that result in the Indian Gaming Regulatory Act." *Seminole Tribe*, 517 U.S. at 75 n.17. Thus, any reluctance on the part of courts to permit actions for injunctions on this ground could be overcome by Congressional authorization of such actions in the context of a particular statute.

[237] *See* Brant, *supra* note 184, at 239. More generally, it is possible that the Court's description in *Seminole Tribe* of *Ex parte Young* as a "narrow" exception to state immunity, *see Seminole Tribe*, 517 U.S. at 76, could herald the doctrine's "potential evisceration." Jackson, *supra* note 236, at 495; *see also* FALLON ET AL., *supra* note 183, at 119. *But see* David P. Currie, Ex parte Young *after* Seminole Tribe, 72 N.Y.U. L. REV. 547, 547 (1997) (arguing that "*Ex parte Young* is alive and well and living in the Supreme Court.") (footnote omitted). In Idaho v. Coeur d'Alene Tribe of Idaho, 521 U.S. 261 (1997), the Supreme Court may have narrowed further the availability of *Ex parte Young* actions. In that case, the Court held that such a remedy is unavailable when the suit threatens a state's sovereignty interests. *See Coeur d'Alene Tribe*, 521 U.S. at 282.

FLSA and FMLA. Finally, recent Supreme Court decisions indicate that *Ex parte Young* actions may be difficult to maintain.

D. Suits Against State Officials

The final enforcement alternative currently available is for state employees to file suits for damages against state officials in their individual capacities. State officials are not protected by state sovereign immunity.[238]

One problem with such an approach is that the federal statute at issue must allow for suits against individual officials. The FLSA defines "employer" as "includ[ing] any person acting directly or indirectly in the interest of an employer in relation to an employee,"[239] and the FMLA has an almost identical definition.[240] Applying these definitions, some courts have held individuals liable under these statutes.[241]

Even if individual officers are liable under the FLSA and FMLA, courts have rejected this approach for the ADEA and ADA.[242] Therefore, in order for suits against state officials to proceed, Congress would likely have to amend the statutory definition of "employer," at least in the ADEA and ADA.[243]

---

[238] *See* Alden v. Maine, 527 U.S. 706, 757 (1999) ("Even a suit for money damages may be prosecuted against a state officer in his individual capacity for unconstitutional or wrongful conduct fairly attributable to the officer himself, so long as the relief is sought not from the state treasury but from the officer personally.") (citing Scheuer v. Rhodes, 416 U.S. 232, 237–38 (1974); Ford Motor Co. v. Department of Treasury of Ind., 323 U.S. 459, 462 (1945)).

[239] 29 U.S.C. § 203(d) (1994).

[240] *See* 29 U.S.C. § 2611(4)(A)(ii)(I) (1994 & Supp. IV 1998) (defining "employer" as including "any person who acts, directly or indirectly, in the interest of an employer to any of the employees of such employer").

[241] *See, e.g.*, United States Dep't of Labor v. Cole Enters., Inc., 62 F.3d 775, 778 (6th Cir. 1995); Donovan v. Agnew, 712 F.2d 1509, 1511 (1st Cir. 1983); Kilvitis v. County of Luzerne, 52 F. Supp. 2d 403, 411–16 (M.D. Pa. 1999). *But see* Wascura v. Carver, 169 F.3d 683, 686 (11th Cir. 1999) (holding that individual government supervisors are not liable under the FLSA or FMLA). Professor Meltzer has suggested that even some of the cases that do allow for suits against individuals "might not support imposing liability, for example, on a middle manager in a large bureaucracy." Meltzer, *supra* note 211, at 1019 n.37.

[242] *See* Butler v. City of Prairie Village, 172 F.3d 736, 744 (10th Cir. 1999) (holding that "the ADA precludes personal capacity suits against individuals who do not otherwise qualify as employers" and noting that "our position [is] consistent with the majority of federal circuit and district courts that have considered the issue of individual supervisor liability under . . . the ADEA . . . [and] with the only circuit courts that have directly addressed the issue of individual liability under the ADA") (citations omitted); *see also* PERRITT, *supra* note 59, § 4.07, at 243–45.

[243] Presumably, state officials who violate the federal employment laws are not susceptible to suit under 42 U.S.C. § 1983 (Supp. IV 1998) because the FLSA, ADEA, ADA, and FMLA likely all contain "comprehensive enforcement mechanisms" that foreclose the § 1983 remedy. Middlesex County Sewerage Auth. v. Nat'l Sea Clammers Ass'n, 453 U.S. 1, 20 (1981); *see also* RICHARD H. FALLON ET AL., HART AND WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 1134–35 (4th ed. 1996); 1 CRAIG, *supra* note 136, §§ 7.5, 7.6, at 530–34. Indeed, many courts appear to have reached this conclusion with regard to the federal employment laws. *See, e.g.*, Kendall v. City of Chesapeake, 174 F.3d 437, 442 (4th Cir. 1999) (FLSA); Zombro v. Baltimore City Police Dep't, 868 F.2d 1364,

Another barrier to suits against individual state officials is that many will be able to assert a qualified immunity defense.[244] Qualified immunity is an "entitlement not to stand trial or face the other burdens of litigation,"[245] and is available to state officials who do not "violate clearly established statutory or constitutional rights of which a reasonable person would have known."[246] One commentator has noted that "seldom will a rule from one case apply with 'obvious clarity' to the facts of another case," making the standard for rejection of the defense a difficult one to meet.[247] Indeed, according to one study, qualified immunity defenses are denied only about twenty percent of the time.[248] Thus, "a court in most cases is able to limit recovery to those plaintiffs it considers to be truly worthy suing state officers the court considers to be truly blameworthy."[249]

Moreover, although suits against individual officials would provide employees with a monetary remedy and might even "pressure states to indemnify their officials,"[250] there could be disadvantages. Specifically, a regime of individual liability for violations of federal employment laws might be considered unfair,[251] or might result in self-protective and risk-minimizing behavior in state officials.[252]

Individual capacity suits against state officials do not provide a means of enforcing federal employment laws equivalent to the private damage suits now prohibited by state sovereign immunity. The individual liability of state officials is uncertain with regard to the FLSA and FMLA, and very unlikely with regard to the ADA and ADEA. Even where individual capacity suits are appropriate, the qualified immunity defense will prevent recovery in many cases.

IV. State Waivers of Immunity

In the final analysis, the most promising enforcement strategy may turn on state waivers of immunity. After all, "a State's sovereign immunity is 'a personal privilege which it may waive at pleasure.'"[253] Thus, if a

---

1366 (4th Cir. 1989) (ADEA); Holbrook v. City of Alpharetta, 112 F.3d 1522, 1531 (11th Cir. 1997) (ADA); *Kilvitis*, 52 F. Supp. 2d. at 416–19 (FMLA).

[244] *See generally* 2 Craig, *supra* note 136, §§ 11.28–11.50, at 151–206.

[245] Mitchell v. Forsyth, 472 U.S. 511, 526 (1985).

[246] Harlow v. Fitzgerald, 457 U.S. 800, 818 (1982).

[247] Hartley, *supra* note 212, at 395 (quoting the phrase "obvious clarity" from United States v. Lanier, 520 U.S. 259, 271 (1997)).

[248] *See* Diana Hassel, *Living a Lie: The Cost of Qualified Immunity*, 64 Mo. L. Rev. 123, 145 n.106 (1999).

[249] Hartley, *supra* note 212, at 399; *see also* Hassel, *supra* note 248, at 147 (arguing that the qualified immunity standard "provides a wide range of discretion and flexibility on the part of the judge").

[250] Meltzer, *supra* note 211, at 1018.

[251] *Cf.* Fallon et al., *supra* note 183, at 149 ("An individual defendant, of course, may not have the resources necessary to satisfy a judgment for substantial damages.").

[252] *See* Peter Schuck, Suing Government 79–81 (1983).

[253] College Sav. Bank v. Florida Prepaid Postsecondary Educ. Expense Bd., 527 U.S.

state chooses to waive its immunity from suits brought by its employees for violations of federal employment statutes, the problem is solved.

State waiver of immunity must be "altogether voluntary on the part of the sovereignty."[254] To this end, it requires a "'clear declaration' that [a state] intends to submit itself to [the federal courts'] jurisdiction."[255] The Supreme Court in *College Savings Bank v. Florida Prepaid Postsecondary Education Expense Board* held that a state's mere participation in a regulated industry cannot constitute a waiver of immunity.[256]

Of course, states will naturally be reluctant to increase their own liability by waiving immunity. This Part discusses three ways that states might be persuaded to do so: through the congressional spending power, through collective bargaining by state employees, or through other methods resulting in unilateral action by a state.

A. Congressional Spending Power

The most likely path toward state waivers of immunity is the congressional spending power. The Constitution gives Congress power to "lay and collect Taxes, Duties, Imposts, and Excises, to pay the Debts and provide for the common Defence and general Welfare of the United States."[257] Pursuant to this power, Congress may "further broad policy objectives by conditioning receipt of federal moneys upon compliance by the recipient with federal statutory and administrative directives."[258]

The Supreme Court explained the constitutional requirements for conditional spending in *South Dakota v. Dole*.[259] In *Dole*, Congress had conditioned five percent of a federal highway grant on the condition that states receiving the funds raise their minimum drinking ages to twenty-one years.[260] The Court upheld this condition as constitutional even though Congress might not have the power to impose a national drinking age directly.[261] It stated that "objectives not thought to be within Article I's 'enumerated legislative fields' may nevertheless be attained through

---

666, 675 (1999) (quoting Clark v. Barnard, 108 U.S. 436, 447 (1883)).

[254] *Id.* (quoting Beers v. Arkansas, 61 U.S. (20 How.) 527, 529 (1858)) (internal quotation marks omitted).

[255] *Id.* at 676 (quoting Great N. Life Ins. Co. v. Read, 322 U.S. 47, 54 (1944)). One commentator has recommended "a thorough state-by-state analysis" to determine whether any states have already waived their immunity for federal claims, Hartley, *supra* note 212, at 368, but given the narrowness with which state immunity waivers are interpreted, *see College Sav.*, 527 U.S. at 676, it seems unlikely that any have done so.

[256] 527 U.S. 666, 687 (1999).

[257] U.S. CONST. art. I, § 8, cl. 1.

[258] South Dakota v. Dole, 483 U.S. 203, 206 (1987) (quoting Fullilove v. Klutznick, 448 U.S. 448, 474 (1980)).

[259] *Id.*

[260] *Id.* at 205, 211.

[261] *Id.* at 212.

the use of the spending power and the conditional grant of federal funds."[262]

In reaching this decision, the Court noted that Congress' spending power is "not unlimited."[263] First, it can be used only to further the "general welfare."[264] Second, conditions must be "unambiguous[ ]" so that states can "exercise their choice knowingly, cognizant of the consequences of their participation."[265] Third, they must be related "to the federal interest in particular national projects or programs."[266] Fourth, there can be no other constitutional provisions that "provide an independent bar to the conditional grant of federal funds."[267]

In *College Savings Bank v. Florida Prepaid Postsecondary Education Expense Board*,[268] the Supreme Court reaffirmed this use of the spending power.[269] Indeed, the *College Savings* Court specifically distinguished the constructive waiver at issue in that case from conditional spending under *Dole*.[270] Justice Scalia's opinion for the Court stated that:

> Congress has no obligation to use its Spending Clause power to disburse funds to the States; such funds are gifts. In the present case, however, what Congress threatens if the State refuses to

---

[262] *Id.* at 207 (quoting United States v. Butler, 297 U.S. 1, 65 (1936)).

[263] *Id.*

[264] *Id.* Because "[i]n considering whether a particular expenditure is intended to serve general public purposes, courts should defer substantially to the judgment of Congress," *see id.* (citing Helvering v. Davis, 301 U.S. 619, 640, 645 (1937)), the "general welfare" requirement is easily met. *See id.* at 207 n.2 ("The level of deference to the congressional decision is such that the Court has more recently questioned whether 'general welfare' is a judicially enforceable restriction at all.") (citing Buckley v. Valeo, 424 U.S. 1, 90–91 (1976)).

[265] *Id.* (quoting Pennhurst State Sch. & Hosp. v. Halderman, 451 U.S. 1, 17 (1981)) (internal quotation marks omitted).

[266] *Id.* (quoting Massachusetts v. United States, 435 U.S. 444, 461 (1978)) (internal quotation marks omitted). In Bradley v. Arkansas Dep't of Educ., 189 F.3d 745 (8th Cir. 1999), the Eighth Circuit held that in accepting federal funds, the states did not consent to be sued for violations of § 504 of the Rehabilitation Act. *Id.* at 758. The Rehabilitation Act prohibits programs and activities that receive federal funds from discriminating on the basis of disability. *See* 29 U.S.C. § 794 (1994 & Supp. IV 1998). In *Bradley*, the court held that the Rehabilitation Act "amounts to impermissible coercion: Arkansas is forced to renounce all federal funding, included funding wholly unrelated to the [Rehabilitation Act], if it does not want to comply with § 504." 189 F.3d at 757. In the same case, however, the court held that federal funds were effectively conditioned on state waiver of immunity for suits under the Individuals with Disabilities Education Act ("IDEA"), 20 U.S.C. §§ 1400–1487 (Supp. V 1999), in which only funding for IDEA programs was conditioned upon waiver of immunity. *Bradley*, 189 F.3d at 753.

[267] Dole, 483 U.S. at 208 (citations omitted). The *Dole* Court interpreted this requirement as "stand[ing] for the unexceptionable proposition that the power may not be used to induce the States to engage in activities that would themselves be unconstitutional." *Id.* at 210.

[268] 527 U.S. 666 (1999).

[269] *Id.* at 686–87. The Supreme Court has never invalidated a conditional offer of funds. *See* Lynn A. Baker, *Conditional Federal Spending After* Lopez, 95 COLUM. L. REV. 1911, 1924 (1995).

[270] *Coll. Sav. Bank*, 527 U.S. at 686–87.

> agree to its condition is not a denial of a gift or gratuity but a sanction: exclusion of the State from otherwise permissible activity.[271]

It is possible that *College Savings* foreshadows the Court's intention to "scrutinize conditional spending programs with renewed vigor."[272] Such a reading of *College Savings* is probably not warranted, however.[273] Instead, it is best to read *College Savings* "as simply reiterating the lenient constitutional standard for conditions attached to spending regimes as enunciated in *Dole*."[274]

In any case, it would not be difficult for an interested Congress, through its spending power, to encourage states to waive their immunity.[275] Immunity waivers could be attached as conditions to any number of federal spending programs that would satisfy the *Dole* germaneness requirement. For example, the federal government gives a significant amount of money to states to support employment programs.[276] It does

---

[271] *Id.* In dissent, Justice Breyer objected that "[g]iven the amount of money at stake, it may be harder, not easier, for a State to refuse highway funds than to refrain from entering the investment services business." *Id.* at 697 (Breyer, J., dissenting).

[272] Gordon L. Hamrick, IV, Comment, *Roving Federalism: Waiver Doctrine After* College Savings Bank v. Florida Prepaid Postsecondary Education Expense Board, 49 EMORY L.J. 859, 890 (2000).

[273] *See id.* at 894–97.

[274] *Id.* at 898 (footnote omitted). In *College Savings*, the Court stated that "[t]here is a fundamental difference between a State's expressing unequivocally that it waives its immunity, and Congress' expressing unequivocally its intention that if the State takes certain action it shall be deemed to have waived that immunity." 527 U.S. at 680–81. This language could be read to "indicate that if a court is to find a waiver of state immunity, the state not only must accept the federal funds involved and participate in the particular spending program, but it also must have enacted a statute specifically waiving its immunity under that program." Hamrick, *supra* note 272, at 889 n.183. Such a cumbersome procedure would not be constitutionally required. The question of who has the authority to waive a state's sovereign immunity is a matter of state law. Ford Motor Co. v. Dep't of Treasury of Ind., 323 U.S. 459, 467 (1945); 2 CRAIG, *supra* note 136, § 10.7, at 68. Presumably, states could require an act of the legislature to waive immunity, but if receipt of certain federal funds is validly conditioned on a waiver, then those funds could not be obtained except by statutory authorization. Such a requirement "would only increase the decision costs accrued by a state when taking steps to qualify for federal dollars." Hamrick, *supra* note 272, at 889 n.183.

[275] Presumably, a state could terminate its immunity waiver at any time by declining the federal funds at issue. *See* Hamrick, supra note 272, at 891 n.190. Indeed, the Supreme Court has held that a state can reverse its decision to waive immunity even in the context of a pending lawsuit. *See* Beers v. Arkansas, 61 U.S. (20 How.) 527, 529 (1857). Still, "there is the empirical question of whether a state actually would choose to terminate a substantial, popular program." Hamrick, *supra* note 272, at 891 n.190.

[276] Among these programs in the 2001 appropriations was funding for State Unemployment Insurance and Employment Service Operations, *see* HOUSE COMM. ON APPROPRIATIONS, DEP'TS OF LABOR, HEALTH AND HUMAN SERVS., AND EDUC., AND RELATED AGENCIES APPROPRIATIONS BILL, 2001, H.R. REP. NO. 106-645, at 17 (2000) (providing "administrative grants and assistance to State agencies which administer Federal and State unemployment compensation laws and *operate the public employment service*") (emphasis added).

not seem unreasonable to require that states receiving federal employment dollars allow their own employees the full benefit of federal employment laws. Also, there are many federal grants that fund the hiring or training of public employees.[277] Certainly, when federal money is used to hire state workers, Congress should be able to require that those workers have the capacity to enforce their federal employment rights effectively.[278]

As was the case in *Dole*, Congress could provide that states that refused to waive their immunity would forfeit a small portion of the funds from the federal grant or grants at issue, rather than the entire amount. This would ensure that the states' acceptance of the conditions is not coerced. States would then weigh the expected benefit from the federal grant against the expected cost of being subjected to federal lawsuits. This expected cost, however, is not simply the anticipated liability from federal suits. As noted above, to some extent, many states already pay damages for employment violations. Indeed, it may be less expensive to be sued in a federal court than in a state forum, in that the federal government would pay administrative costs. Thus, states may be willing to waive their sovereign immunity from employment suits even if the likely judgments against them are equal to, or slightly more than, the amount to be gained in grants.

Conditioning spending on state waivers of immunity could prove difficult politically, but it is not implausible. After all, Congress attaches waivers of various kinds to many of the grants it makes to states.[279] Moreover, Congress passed the FLSA, ADEA, ADA, and FMLA, and voted to include state employees in the coverage of each statute.[280] Justice Breyer may be correct that the Supreme Court is "simply impos[ing] upon Congress the burden of rewriting legislation, for no apparent rea-

---

[277] Federal money directly or indirectly supports the employment of educators through Eisenhower Professional Development State Grants, *see id.* at 138, and Teacher Quality Enhancement Grants, *see id.* at 167–68; corrections officers through State Prison Grants, *see* HOUSE COMM. ON APPROPRIATIONS, DEP'TS OF COMMERCE, JUSTICE, AND STATE, THE JUDICIARY, AND RELATED AGENCIES APPROPRIATIONS BILL, 2001, H.R. REP. NO. 106-680, at 47 (2000); police officers through the Edward Byrne Memorial State and Local Law Enforcement Assistance Program, *see id.* at 48, and Community Oriented Policing Services Police Hiring Initiatives, *see id.* at 53; judges, probation officers, and lawyers through Juvenile Accountability Incentive Block Grants, *see id.* at 49; and highway and mass-transit workers through federal highway and mass-transit grants, *see* HOUSE COMM. ON APPROPRIATIONS, DEPARTMENT OF TRANSP. AND RELATED AGENCIES APPROPRIATIONS BILL, 2001, H.R. REP. NO. 106-622, at 66, 103–13 (2000). This is to say nothing of the numerous grants provided to individual state agencies and programs.

[278] Dole, 483 U.S. 203, 205, 211 (1987).

[279] *See* Baker, *supra* note 269, at 1918 (noting that federal funds make up a large proportion of state revenues and stating that "none of this federal money is offered the states unconditionally").

[280] *See* Fair Labor Standards Amendments of 1985, Pub. L. No. 99-150, 99 Stat. 787 (1985) (codified at 29 U.S.C. § 207(o) (1994 & Supp. IV 1998)); Fair Labor Standards Amendments of 1974, Pub. L. No. 93-259, 88 Stat. 55 (1974) (codified at 29 U.S.C. § 630(b) (1994)); 42 U.S.C. §§ 12111(5), 12202 (1994); 29 U.S.C. § 2617(a) (1994).

son" to condition spending on immunity waivers,[281] but there is no real reason to believe that Congress would be unwilling to undertake this burden, at least to restore the status quo ante.

B. Collective Bargaining

Waiver might also be accomplished through bargaining by the state employees themselves. Collective bargaining offers a possible solution to the problem faced by public employees. Unions may be able to negotiate for a waiver of sovereign immunity or for equivalent protections through the grievance arbitration process.

Over 1.8 million state employees are union members.[282] Indeed, "[p]ublic sector collective bargaining is one of the few areas in which unions can still claim some degree of growth."[283]

Importantly, however, not every state has a strong union presence. Public employers are exempted from the coverage of the National Labor Relations Act.[284] For this reason, collective bargaining rights for state employees can come only from state statutes.[285] Twenty-seven states guarantee some form of collective bargaining rights to state employees.[286] These rights, however, vary greatly from state to state.[287] Twenty-three states do not allow any collective bargaining by state employees at all.[288]

Where state workers have collective bargaining rights, over half of state employees are unionized, but where they do not enjoy such rights, only slightly over ten percent have joined unions.[289] Thus, to the extent that union activity provides a solution to state employees' difficulty in enforcing federal employment rights, it is likely to succeed only in states that permit collective bargaining.

---

[281] *Coll. Sav. Bank*, 527 U.S. at 704 (Breyer, J., dissenting).

[282] *See* PUBLIC EMPLOYEE DEPARTMENT, AFL-CIO, PUBLIC EMPLOYEES BARGAIN FOR EXCELLENCE: A COMPENDIUM OF STATE PUBLIC SECTOR LABOR RELATIONS LAWS 4 (1995) [hereinafter BARGAIN FOR EXCELLENCE] (using 1992–1994 averaged figures).

[283] DILTS ET AL., *supra* note 19, at 3.

[284] 29 U.S.C. § 152(2) (1994) ("The term 'employer' . . . shall not include . . . any State or political subdivision thereof . . . ."); *see also* EDWARDS ET AL., *supra* note 25, at 85. Bills are frequently introduced in Congress which would subject public employers to federal labor laws, *see id.* at 87–98, but none have been successful.

[285] *See* DILTS ET AL., *supra* note 19, at 17; EDWARDS ET AL., *supra* note 25, at 139–46.

[286] *See* BARGAIN FOR EXCELLENCE, *supra* note 282, at 1. These states are Alaska, California, Connecticut, Delaware, Florida, Hawaii, Illinois, Indiana, Iowa, Kansas, Maine, Massachusetts, Michigan, Minnesota, Montana, Nebraska, New Hampshire, New Jersey, New Mexico, New York, Ohio, Oregon, Pennsylvania, Rhode Island, Vermont, Washington, and Wisconsin. *See id.* at 3–4.

[287] *See generally id.* at 8–56 (detailing the provisions of each state's collective bargaining laws); *see also* DILTS ET AL., *supra* note 19, at 19–23.

[288] These states are Alabama, Arizona, Arkansas, Colorado, Georgia, Idaho, Kentucky, Louisiana, Maryland, Mississippi, Missouri, Nevada, North Carolina, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, West Virginia, and Wyoming. *See* BARGAIN FOR EXCELLENCE, *supra* note 282, at 3–4.

[289] *See id.*

Even in states that do allow state employees to bargain collectively, public employee unions lack many of the tools that private sector unions have at their disposal. For example, some states only allow collective bargaining over certain working conditions.[290] Many states do not grant public employees the right to strike,[291] depriving state workers of an important "economic weapon."[292]

Still, it is possible that in some states, state employees may be able to bargain for a state waiver of immunity. Again, there is an issue as to whether the state's bargaining representative could enter into an agreement with employees to waive immunity. If a waiver could only be accomplished through statute, for example, then the collective bargaining alternative would likely not succeed.[293]

A related possibility is for state employee unions to bargain for the incorporation of the federal employment laws into their collective bargaining agreements. This is apparently a course of action advocated by the American Federation of State, County, and Municipal Employees ("AFSCME").[294] With this strategy, employees could enforce their federal employment rights through the grievance arbitration process.[295]

There are several disadvantages to the arbitration approach. First, there are significant differences between arbitration and court proceedings.[296] Second, arbitrators are not bound by stare decisis,[297] so there is no guarantee that an arbitrator's interpretation of a federal employment stat-

---

[290] *See id.* at 1.

[291] *See* DILTS ET AL., *supra* note 19, at 21 (noting that only nine states allow even a limited right to strike); *see generally* BARGAIN FOR EXCELLENCE, *supra* note 282. Even states that do allow a right to strike often only allow work stoppages in some circumstances. *See* EDWARDS ET AL., *supra* note 25, at 711–12.

[292] *Cf.* Paul Weiler, *Striking a New Balance: Freedom of Contract and the Prospects for Union Representation*, 98 HARV. L. REV. 351, 386 (1984) (discussing "economic weapons").

[293] *Cf. supra* note 274 (discussing who can validly waive a state's sovereign immunity).

[294] *See Supreme Court vs. State Employees*, COLLECTIVE BARGAINING REPORTER (2000), *available at* http://www.afscme.org/wrkplace/cbr100%5F2.htm (last visited Oct. 3, 2001). AFSCME proposes negotiating contract language similar to the following:

> The parties recognize the applicability of the Family and Medical Leave Act (FMLA), the Americans with Disabilities Act (ADA), the Equal Pay Act (EPA), the Age Discrimination in Employment Act (ADEA), and the Fair Labor Standards Act (FLSA) to bargaining unit employees. Complaints concerning any claimed violation, misinterpretation, or misapplication of such Federal statutes, or any applicable rules and regulations implementing such statutes, may be filed as a grievance and pursued to arbitration in accordance with the terms of this Agreement. The arbitrator shall have authority to remedy any violation to the extent consistent with the applicable Federal law.

Id.

[295] *See id.*

[296] *See generally* DEITSCH & DILTS, *supra* note 21, at 55–72 (discussing procedure and evidence rules in arbitration proceedings).

[297] *See id.* at 81–82.

ute will be consistent with that arrived at by courts. Finally, it is possible that employees will have difficulty enforcing arbitration awards against their state employers if a state claims it is entitled to sovereign immunity.[298]

Resolution of federal employment rights through the grievance arbitration system thus has significant disadvantages, and may be no better than adjudication of such claims in the diverse state forums. In some cases, however, state employees may prefer the arbitration of federal claims to the adjudication of state claims in state courts or administrative forums. In any case, arbitration issues are not unique to the state employment context. Many employment contracts today provide for the submission of employment-related claims to binding arbitration, a practice recently upheld by the Supreme Court.[299]

C. *Unilateral Waivers*

It is possible, though unlikely, that states will decide unilaterally to waive their sovereign immunity from federal employment suits. After all, as noted above, every state prohibits age and disability discrimination, and many states regulate fair labor standards and provide for family and medical leave.[300] Conceivably, some of these states might not be opposed to allowing their employees to file suits against them for violations of federal laws.

At least fourteen states opposed the Supreme Court's decision in *Garrett*. These states submitted an *amicus* brief arguing against the position that state officials "miraculously were not subject to the same cultural biases and irrational fears about individuals with disabilities" as private employers.[301] Therefore, at least with regard to the ADA, there may be support among the states for unilateral waivers of sovereign immunity.[302]

---

[298] Presumably, however, state labor laws allowing for collective bargaining and the enforcement of arbitration awards, *see* EDWARDS ET AL., *supra* note 25, at 796–98, would preclude state governments from raising an immunity defense if a successful grievant sought to collect against an unwilling state employer. *Cf.* Williams v. Lane, 818 F. Supp. 1212, 1213 (N.D. Ill. 1993) (holding that a state, by its entry into a settlement agreement, waived its immunity from suit for violation of the agreement).

[299] *See* Circuit City Stores v. Adams, 532 U.S. 105 (2001) (holding an employee's employment discrimination claim subject to arbitration where the employee had signed an agreement to arbitrate as a condition of his employment).

[300] *See supra* Part III.A.

[301] Editorial, *Backsliding on Rights*, BOSTON GLOBE, Feb. 24, 2001, at A14.

[302] Of course, even those states that opposed the decision in *Garrett* may be reluctant to give up their new-found immunity. When I asked New York Attorney General Eliot Spitzer, who participated in the *amicus* brief, whether his state should waive its sovereign immunity for ADA claims brought by state employees, he said that it should not, citing his obligation to assert all available defenses. Eliot Spitzer, How I Learned to Love the New Federalism, Remarks at Harvard Law School Forum (Mar. 7, 2001). Spitzer acknowledged that the state legislature could choose to allow such suits, but he did not express an opinion

Additionally, states could be pressured to waive their immunity through the political process. If enough voters are unhappy about the current state of the law regarding state employee rights, they may lobby their legislatures to waive immunity.

Whether they are accomplished by spending incentives, bargaining, or political pressure, waivers of immunity by states appear to be the best way for state employees to enforce their rights under federal employment laws. The next Part examines several other less promising possibilities.

V. FEDERAL LEGISLATIVE SOLUTIONS

Of the possible enforcement alternatives and strategies discussed above, two would require some measure of congressional action. First, Congress could dramatically increase the funding it devotes to federal enforcement of employment rights. As noted above, this is probably unlikely from a political perspective. Second, Congress could condition a portion of federal funding to states on waivers of immunity.

This Part examines two other federal legislative solutions: valid abrogation of immunity by Congress and *qui tam* actions. Both of these options, however, are less preferable than state waiver of immunity for two reasons. First, these strategies would require significantly more congressional action than state waivers, including waivers upon which federal funds were conditioned. Second, they are less likely to address the problem adequately and to provide state employees with a way to enforce their federal employment rights effectively.

For any of these four legislative solutions to occur, Congress would need to be motivated to change the law. As noted above, there is reason to believe that Congress might be interested in restoring the status quo ante of subjecting states to suit for federal employment law violations.[303]

A. Valid Abrogation

Congress could attempt to abrogate state immunity pursuant to § 5 of the Fourteenth Amendment.[304] The Supreme Court's decisions in *Kimel* and *Garrett*, however, suggest that this would be an uphill battle. Still, the language in those cases does not indicate that valid abrogation for federal employment laws is impossible.

The Court's view that Congress had not acted on the basis of sufficient evidence of unconstitutional discrimination by state employers

---

on whether it should do so. *Id.*

[303] As Representative Bob Barr (R-Ga.) has said, "[y]ou can never really predict what Congress is going to do." Bob Barr, Remarks at Symposium Presented by the Harvard Journal on Legislation on Reforming Punitive Damages (Mar. 13, 2001).

[304] *Cf. supra* note 124.

was important to its decisions in both *Kimel*[305] and *Garrett*.[306] It is therefore possible that if Congress assembled more evidence of unconstitutional discrimination by states, it could validly abrogate state sovereign immunity.

This would be no easy task, however. Justice Breyer's dissenting opinion in *Garrett* argued that Congress used "hundreds of examples" that reasonably led to a belief that "these examples represented signs of a widespread problem of unconstitutional discrimination."[307] Whether or not this evidence should have been sufficient to justify the statute,[308] it shows that if Congress wants to abrogate state sovereign immunity validly, it will need to gather much more supporting evidence that details a pattern of unconstitutional discrimination by state governments, rather than isolated incidents by private actors.[309]

Assuming that such evidence exists, a Congress that wanted to make use of it to abrogate state immunity would probably need to hold extensive hearings. Conceivably, sufficient evidence of state discrimination could justify abrogation of state immunity. Yet, given the time and effort this would take, coupled with such a strategy's uncertain chances for success, it would probably not be the most effective course of action.

B. Qui tam *Actions*

Another possible legislative solution would be congressional authorization of *qui tam* actions by state employees. In a *qui tam* action, a private individual files suit on her own behalf, but also on behalf of the

---

[305] *See* Kimel v. Florida Bd. of Regents, 528 U.S. 62, 89 (2000) ("Congress never identified any patttern of age discrimination by the States, much less any discrimination whatsoever that rose to the level of constitutional violation."); *id.* at 91 ("Congress had virtually no reason to believe that state and local governments were unconstitutionally discriminating against their employees on the basis of age.").

[306] *See* Board of Trs. of Univ. of Ala. v. Garrett, 531 U.S. 356, 357 (2001) ("[T]hese incidents [of state discrimination] taken together fall far short of even suggesting the pattern of unconstitutional discrimination on which § 5 legislation must be based."); *id.* at 370 ("It is telling, we think, that . . . Congress assembled only such minimal evidence of unconstitutional state discrimination in employment against the disabled."); *id.* at 376 (Kennedy, J., concurring) ("The predicate for money damages against an unconsenting State in suits brought by private persons must be a federal statute enacted upon the documentation of patterns of constitutional violations committed by the State in its official capacity.").

[307] *Id.* at 382 (Breyer, J., dissenting); *see also id.* at 389–424 (listing evidence relied upon by Congress).

[308] *Cf. id.* at 370 (labeling Justice Breyer's examples "unexamined, anecdotal accounts").

[309] *See Kimel*, 528 U.S. at 89; *Garrett*, 531 U.S. at 370–72. For an examination of the impact of state policies on employees and applicants with disabilities, see generally Charles Everet Drum, State Employment Policies: Assessing the Impacts on People with Significant Disabilities (1994) (unpublished dissertation, Brandeis University); *id.* at Abstract (noting "the presence of numerous procedural barriers to the establishment of supported employment programs and the employment of people with significant disabilities").

government.[310] The most frequently used federal statute providing for *qui tam* actions is the False Claims Act ("FCA").[311] The FCA allows a "relator" to bring a *qui tam* action "for the person and for the United States Government" against a person who makes a false claim to the United States.[312]

Since states are not immune from suit by the federal government, this exception may "also encompass[ ] suits that seek to vindicate identical interests of the United States, but are commenced and prosecuted by private individuals rather than by public officials."[313] Courts have split on this issue,[314] and the Supreme Court recently declined to reach the question.[315]

Language in *Alden*[316] and the Court's recent trend in state immunity cases suggest that given the opportunity, it is quite possible that the Court would hold *qui tam* actions against states to be barred by state immunity. Such an outcome, however, would be contrary to the Court's frequent emphasis on originalism.[317] Indeed, public enforcement by private individuals has a long history in the United States, dating back to before the time of the framing of the Constitution.[318]

Even if the Court were to uphold *qui tam* suits in the context of the FCA,[319] it may be reluctant to do so if Congress were to amend the federal employment laws to provide for such a mechanism. The central issue in determining whether immunity applies to a *qui tam* action "is whether the United States' interests are truly at stake in a suit against a state."[320]

---

[310] "*Qui tam*" comes from the Latin "*Qui tam pro domino rege quam pro se ipso in hac parte sequitur*," meaning "who as well for the king as for himself sues in this matter." BLACK'S LAW DICTIONARY 1262 (7th ed. 1999).

[311] 31 U.S.C. §§ 3729–3733 (1994).

[312] *Id.* § 3730(b)(1). The FCA also allows the federal government to bring such a claim itself. *See id.* § 3730(a). Even when private citizens bring suit, the DOJ retains significant control over the litigation and has the power to intervene. *See id.* § 3730.

[313] Evan H. Caminker, *State Immunity Waivers for Suits by the United States*, 98 MICH. L. REV. 92, 98 (1999).

[314] *See id.* at 94 & n.14 (citing cases).

[315] *See* Vermont Agency of Natural Res. v. United States *ex rel.* Stevens, 529 U.S. 765, 773 n.4 (2000) (noting that the Court "leave[s] open today" whether the Eleventh Amendment bars *qui tam* actions against nonconsenting states).

[316] In *Alden*, the Court stated that "[a] suit which is commenced and prosecuted against a State in the name of the United States by those who are entrusted with the constitutional duty to 'take Care that the Laws be faithfully executed' differs in kind from the suit of an individual." Alden v. Maine, 527 U.S. 706, 755 (1999) (quoting U.S. CONST. art. II, § 3).

[317] *See* Caminker, *supra* note 313, at 126–32 (arguing that "states would have assumed that their immunity waiver for suits by the United States encompassed *qui tam* suits on behalf of the United States").

[318] *See id.* Professor Caminker's argument that an originalist interpretation supports allowing *qui tam* actions against states is perhaps bolstered by the Court's discussion in *Stevens* of the historical basis for such suits. *See* Stevens, 529 U.S. at 774–78.

[319] In *Stevens*, the Court interpreted the language of the FCA to preclude suits against states by holding that a state is not a "person" within the meaning of the statute. *See* Stevens, 529 U.S. at 787.

[320] Caminker, *supra* note 313, at 113.

This standard may be easier to meet for the FCA than for the federal employment laws.

In FCA cases, the proprietary interests of the federal government are clearly at stake, because the allegation is that someone has defrauded the United States.[321] In contrast, the federal interest in an individual employee's claim under the FLSA, ADEA, ADA, or FMLA is less concrete.[322] Also, in FCA cases, the relator only receives a percentage of the recovery,[323] but in employment suits, all the recovery would go to the employee.

For these reasons, if Congress were to amend the federal employment laws to provide for *qui tam* suits against states, the Court might "suspiciously view it as an effort to circumvent the *Seminole Tribe/Alden* rule that Congress cannot authorize private parties to assert their 'own' interests against states."[324]

Nevertheless, if *qui tam* actions against states are constitutional, then they should also be allowed for federal employment claims. It is true that the federal government has no financial stake in employment cases, but it does have a significant interest in ensuring that all employers comply with federal law. Indeed, it is because of this federal interest that Congress gave federal agencies the authority to enforce these laws themselves.

Also, just as in FCA cases, the federal government in employment cases retains some level of control over the litigation. The government can preempt or intervene in employee suits brought under the FLSA,[325] ADEA,[326] ADA,[327] and FMLA.[328]

---

[321] *See id.* at 133.

[322] The federal interest in such suits is "the United States' general regulatory interest in enforcing" the federal employment laws. *Id.* at 117.

[323] *See id.* at 101 (noting that successful FCA plaintiffs generally receive between ten and thirty percent of the recovery, depending on the circumstances of the case).

[324] *See* Caminker, *supra* note 313, at 134. Professor Caminker notes a concern about whether Congress can be seen as "cleverly 'delegating back' to a private party authority to bring a claim that was already hers, albeit in a form barred by state sovereign immunity." *Id.* On the other hand, Justice Stevens, dissenting in *Stevens*, wrote that "[i]t is not at all clear to me . . . why a *qui tam* action would be considered an 'end run' around [the Eleventh] Amendment, yet precisely the same form of action is not an 'end run' around Articles II and III." Vermont Agency of Natural Res. v. United States *ex rel.* Stevens, 529 U.S. 765, 802 n.12 (2000) (Stevens, J., dissenting).

[325] *See* 29 U.S.C. § 216(b) (1994) ("The right provided by this subsection to bring an action by or on behalf of any employee, and the right of any employee to become a party plaintiff to such an action, shall terminate upon the filing of a complaint by the Secretary of Labor . . . .").

[326] *See* 29 U.S.C. § 626(c)(1) (1994) ("[T]he right of any person to bring such action shall terminate upon the commencement of an action by the Equal Employment Opportunity Commission to enforce the right of such employee under this [Act].").

[327] *See* 42 U.S.C. § 2000e-5(f)(1) (1994) ("Upon timely application, the court may, in its discretion, permit the Commission, or the Attorney General in a case involving a government, governmental agency, or political subdivision, to intervene in such civil action upon certification that the case is of general public importance.").

[328] *See* 29 U.S.C. § 2617(a)(4) (1994) ("The right . . . to bring an action by or on behalf

Thus, *qui tam* actions may provide a way for individual employees to sue state employers for violations of their federal employment rights. Of course, in order for such an option to be available, Congress would have to amend each of the federal employment laws to provide for such actions. Because of questions about their constitutionality and the significant amount of congressional action this change would require, providing *qui tam* actions may not be the best way of solving the problem. It is, however, still a possible strategy.

VI. CONCLUSION

Federal employment laws create rights for employees and impose obligations on employers. Congress has the power to regulate state employers when it enacts such laws, and it did so in the FLSA, ADEA, ADA, and FMLA. In its recent state sovereign immunity decisions, however, the Supreme Court has taken from state employees the ability to sue their employers for damages, eliminating employees' most effective means of enforcing their employment rights and obtaining adequate remedies.

The remaining enforcement alternatives are ineffective or inadequate. State employment laws vary widely, and many do not provide for substantive rights or remedies that are equivalent to those found in the federal laws. Many do not even provide for a private right of action. Enforcement of employment laws by the federal government is unrealistic because of limited government resources and a large number of cases. Suits for injunctive relief, when available at all, cannot provide injured employees with damages to compensate them for their injuries. Actions against state officials individually, even when permitted by statutory language, will often be barred by the qualified immunity doctrine.

If federal employment laws are to be applied meaningfully to state employers, state employees must have an effective way to enforce their rights. Without significant changes in currently available enforcement alternatives, such as dramatically increased funding for federal enforcement, no such effective method is available. For this reason, a Congress concerned about the rights of all employees must act to ensure that state employers comply with the law.

The most promising enforcement strategy involves state waivers of immunity. This could be accomplished in several ways, principally by conditioning certain federal grants to states on immunity waivers. Also, public employee unions may be able to bargain with state employers for such a waiver or its equivalent, and a few states may be willing to waive their sovereign immunity unilaterally.

---

of any employee shall terminate . . . on the filing of a complaint by the Secretary . . . .").

Other legislative solutions include valid abrogation of state immunity and congressional authorization of *qui tam* lawsuits against states. Recent cases, however, indicate that the former would be difficult to do properly. The latter might avoid the problem, but is itself of uncertain constitutionality.

However appropriate the Supreme Court's recent pronouncements on the issue, state sovereign immunity is the law today. Accordingly, attention should be focused on enforcement alternatives and strategies for state employees, such as those discussed in this Note, so that federal employment rights do not become meaningless for want of effective remedies.

RECENT DEVELOPMENTS

THE UNBORN VICTIMS OF VIOLENCE ACT

On April 26, 2001,[1] the House of Representatives passed the Unborn Victims of Violence Act ("UVV")[2]; the bill currently awaits a vote in the Senate.[3] The UVV amends Title Eighteen of the United States Code and the Uniform Military Code of Justice to provide that one who injures or kills a fetus during the commission of certain enumerated federal crimes[4] is guilty of an offense separate from that against the pregnant woman, and shall be punished under federal law as if the offense had been committed against a person.[5] The Act exempts physicians and women participating in consensual abortions from its provisions.[6] Most notably, though, the UVV defines a fetus as a "child in utero" or "a member of the species homo sapiens, at any stage of development, who is carried in the womb."[7] If enacted as law, the UVV would be the first federal statute to recognize the unborn as independent crime victims equivalent to born persons.[8]

Although the law tends to increase penalties for assaulting a pregnant women by punishing harm done to her fetus during the attack, opponents of the UVV fear that the Act will not, in fact, protect women. By giving the fetus human status, the Act will infringe on a woman's constitutional right to choose an abortion as established in *Roe v. Wade*.[9] *Roe* held that the unborn fetus is not a "person"[10] for the purposes of the Fourteenth Amendment's guarantee that no "person" shall be "deprive[d] of life, liberty, or property without due process of law,"[11] and thereby se-

---

[1] 147 CONG. REC. H1649–50 (daily ed. Apr. 26, 2001).

[2] H.R. 503, 107th Cong. (2001). The Unborn Victims of Violence Act was introduced on February 7, 2001 by Rep. Gene Green (R-Tex.), *see* 147 CONG. REC. H222 (daily ed. Feb. 7, 2001), and passed the House by a vote of 252-172. *See* 147 CONG. REC. H1649–50 (daily ed. Apr. 26, 2001). Identical legislation also passed the House in 1999 but died in the Senate under threat of a presidential veto. *See* Aaron Wagner, *Texas Two-Step: Serving up Fetal Rights by Side-Stepping* Roe v. Wade *Has Set the Table for Another Showdown on Fetal Personhood in Texas and Beyond*, 32 TEX. TECH L. REV. 1085, 1087 n.11 (2001).

[3] 147 CONG. REC. S6016 (daily ed. June 8, 2001) (placing H.R. 503 on the calendar in the Senate).

[4] *See* H.R. 503, §§ 2–3. Such crimes include drive-by shootings in furtherance of or to escape a major drug offense, see 18 U.S.C. § 36 (1994); animal enterprise terrorism, see 18 U.S.C. § 43 (1994); and violations of the Freedom of Access to Clinic Entrances Act, see 18 U.S.C. § 248 (1994).

[5] *See* H.R. 503, §§ 2–3. The Act, however, specifically prohibits death sentences for offenders. *Id.*

[6] *See* H.R. 503, §§ 2–3; *see also infra* text accompanying notes 46–49.

[7] H.R. 503, §§ 2–3.

[8] *See* UNBORN VICTIMS OF VIOLENCE ACT OF 2001, H.R. REP. NO. 107-42, pt. 1, at 77–78 (107th Cong., 2001); *see also* Barry J. Lipson, *Federally Speaking*, 3 NO. 11 LAW. J. 7, 18 (2001).

[9] 410 U.S. 113 (1973).

[10] *Id.* at 157.

[11] U.S. CONST. amend XIV, § 1.

cured the right of a pregnant woman to choose to abort a nonviable fetus.[12] In spite of this facial inconsistency with *Roe*, state courts, reasoning from Supreme Court precedent, have upheld fetal homicide laws similar to the UVV because those state statutes specifically exempted consensual abortions.[13] The Supreme Court may view these state court judgments as persuasive if it ultimately rules on the constitutionality of fetal murder statutes. Nevertheless, by treating a fetus as a person for the purposes of federal criminal law, the UVV may lead some to question *Roe*'s assessment of fetal life. Coupled with improvements in prenatal medicine and technology, the Act may in fact serve ultimately to undermine abortion rights.

Statutory criminalization of feticide arose against the background of the common law "born alive" rule,[14] which held that a fetus could not be the subject of homicide until it was born alive and then died of injuries inflicted on it by another person.[15] Under the common law, "homicide" signified the killing of one human being by another.[16] The then-prevailing philosophy held that only a born entity was a "human being."[17] By definition then, a fetus could not be murdered.[18] American common law retained the "born alive" rule until the mid-nineteenth century, when states began to institute criminal statutes against feticide.[19]

The "born alive" rule stemmed primarily from a lack of sophisticated medical knowledge.[20] Miscarriage and infant mortality rates were high at the time of the adoption of the "born alive" rule.[21] Requiring that a fetus be born alive and then die of inflicted injuries overcame the presumption that it would have died or never have been born without the defendant's intervention; thus the rule provides a causal connection between the defendant's conduct and the child's death.[22] Moreover, it was often difficult for a woman to determine whether she was pregnant, as the symptoms of pregnancy matched those of several diseases.[23] Requiring

---

[12] *See Roe*, 410 U.S. at 163.

[13] *See, e.g*, *People v. Davis*, 872 P.2d 591 (Cal. 1994); *State v. Merrill*, 450 N.W.2d 318 (Minn. 1990).

[14] *See Keeler v. Superior Court of Amador County*, 470 P.2d 617, 621 (Cal. 1970) (discussing statutory development of feticide laws in contrast to the common law "born alive" rule).

[15] *See id.* at 620.

[16] *See* Katharine B. Folger, *When Does Life Begin . . . or End? The California Supreme Court Redefines Fetal Murder in* People v. Davis, 29 U.S.F. L. REV. 237, 239 (1994).

[17] *See* Bicka A. Barlow, *Severe Penalties for the Destruction of "Potential Life"—Cruel and Unusual Punishment?*, 29 U.S.F. L. REV. 463, 467 (1995).

[18] *See id.* at 467; *see also* Folger, *supra* note 16, at 241.

[19] *See Keeler*, 470 P.2d at 621.

[20] *See* Barlow, *supra* note 17, at 467.

[21] *See id.*

[22] *See id.* at 468; *see also* Alison Tsao, *Fetal Homicide Laws: Shield Against Domestic Violence or Sword to Pierce Abortion Rights?*, 25 HASTINGS CONST. L.Q. 457, 461 (1998).

[23] *See* Barlow, *supra* note 17, at 467.

live birth for a homicide conviction ensured that no defendant could be convicted of fetal homicide when no fetus in fact existed.[24]

With contemporary medical technology substantially eliminating these problems, the rationale behind the "born alive" rule now seems less compelling, and many states have moved to criminalize feticide.[25] Although the statutory language adopted by each state varies considerably, two basic dividing lines are evident. First, fetal homicide laws select one of several points in gestation after which criminal punishment accrues.[26] States utilize five stages of gestation—birth,[27] viability,[28] quickening,[29] post-embryo,[30] and conception[31]—in determining when the killing of a

---

[24] *See id.*

[25] *See* Tsao, *supra* note 22, at 461.

[26] *See* Barlow, *supra* note 17, at 466.

[27] A statute that makes birth of the victim the point after which penalties for homicide attach is equivalent to the "born alive" rule. Although the "born alive" rule has been replaced with feticide statutes in about half of the states, many states retain a modified version of the rule. *See* Tsao, *supra* note 22, at 461. For states that retain a version of the "born alive" rule, see, for example, *People v. Vercelletto*, 514 N.Y.S.2d 177 (N.Y. Co. Ct. 1987) (holding that, because "person" as statutorily defined for homicide refers only to a born human being, defendant could not be liable for vehicular manslaughter for the death of an unborn fetus).

[28] Although viability has varying definitions, the "generally accepted scientific definition" is the point when "the fetus will have 'a reasonable potential for subsequent survival if . . . removed from the uterus.'" Barlow, *supra* note 17, at 471 (quoting F. GARY CUNNINGHAM ET AL., WILLIAMS OBSTETRICS 501 (18th ed. 1989)). Viability generally occurs between twenty and twenty-four weeks gestation, at about the end of the second trimester of pregnancy. *See id.* For states that adopt viability as the dividing line for criminalizing feticide, see, for example, *State v. Horne*, 319 S.E.2d 703 (S.C. 1984) (holding that an action can be maintained for fetal homicide if "the fetus involved was viable, i.e., able to live separate and apart from its mother without the aid of artificial support"); IOWA CODE ANN. § 707.7 (West 1993 & Supp. 1996); N.Y. PENAL LAW § 125.00 (McKinney 1997).

[29] Quickening is the point at which a fetus first moves within the womb or is capable of moving; this stage usually occurs between the sixteenth and twentieth week of pregnancy. *See* Tsao, *supra* note 22, at 463; *see also* Barlow, *supra* note 17, at 474. For states that criminalize feticide after quickening, see, for example, GA. CODE ANN. § 16-5-80(a) (1992). Georgia criminalizes fetal homicide under a separate feticide statute, which includes only "quick" fetuses: "[a] person commits the offense of feticide if he willfully kills an unborn child so far developed as to be ordinarily called 'quick' by any injury to the mother of such child, which would be murder if it resulted in the death of such mother." *Id. See also* NEV. REV. STAT. § 200.210 (2000); OKLA. STAT. tit. 21 § 713 (1983); WASH. REV. CODE ANN. § 9A.32.060 (West 2000); FLA. STAT. ANN. § 782.09 (West 2000); MICH. COMP. LAWS ANN. § 750.322 (West 1991); MISS. CODE ANN. § 97-3-37 (2000); R.I. GEN. LAWS § 11-23-5 (2000) (defining "quick" under Rhode Island law such that it could be interpreted to mean "viable": "so far developed and matured as to be capable of surviving the trauma of birth with the aid of usual medical care and facilities available in this state.").

[30] The post-embryo stage, when an embryo nominally becomes a fetus, occurs approximately seven to eight weeks into gestation. *See Davis*, 872 P.2d at 599. California is the only state to adopt the post-embryo stage for fetal homicide. *See id.* at 602 (ruling that viability was not a necessary element of fetal murder under California's feticide statute and that the most appropriate definition of "fetus" was the medical one—a post-embryonic unborn child).

[31] For states that criminalize feticide at all points after conception, see, for example, ARIZ. REV. STAT. § 13-1103(A)(5) (2001); ILL. COMP. STAT. Ch. 720 (West 1993); LA. REV. STAT. ANN. § 14:32.5–.8 (West 1987); MINN. STAT. § 609.2661–.2665 (1998); MO. ANN. STAT. §§ 1.205, 565.020, 565.024 (West 1999); N.D. CENT. CODE § 12.1-17.1-01 to

fetus deserves criminal punishment equivalent to that provided for killing born human beings. One state may even choose to criminalize nonconsensual termination of a pregnancy at any point after conception, while not punishing consensual termination until after viability.[32]

The second major dividing line in states' approaches to feticide laws separates states that include fetuses as distinct entities in their criminal codes from those that interpret the words "person" or "human being" in their statutes to include fetuses. California, for example, falls into the former category. After the California Supreme Court held that the use of "human being" in the murder statute did not include unborn children,[33] the state legislature amended the statute to include fetus victims as a separate category.[34] In so doing, the legislature avoided equating a fetus with a full human being.[35] South Carolina case law, by contrast, subsumes fetuses within the classification of "person." The South Carolina Supreme Court determined that a viable fetus could be the subject of homicide.[36] Because the murder statute refers only to a "person" as its subject,[37] the court's statutory interpretation has, in effect, classified viable fetuses as persons.

The UVV would be the first federal law to define a fetus as a "child in utero" or "person" for the purposes of criminal law, and thus would grant "unparalleled federal protection to the human fetus."[38] The UVV imposes penalties on those who injure or kill a fetus in the process of attacking a pregnant woman and, unlike many state statutes,[39] does not require proof that an assailant "had knowledge or should have had knowledge that the victim of the underlying offense was pregnant."[40] This absence of an intent requirement is justified by the theory of "transferred

---

12.1-17.1-04 (1987); OHIO REV. CODE ANN. § 2903.01–.06, 2903.08–.09 (West 1998 & Supp. 2000); 18 PA. CONS. STAT. ANN. §§ 2601–2609 (West 1998); S.D. CODIFIED LAWS §§ 22-16-1, 22-16-1.1, 22-16-15(5), 22-16-20, and 22-16-41 read with §§ 22-1-2(31), 22-1-2(50A) (Michie Supp. 1995 & Supp. 2000); UTAH CODE ANN. §§ 76-5-201 (Supp. 1998); WISC. STAT. ANN. §§ 939.24–.25, 939.75, 940.01–.02, 940.05, 940.08–.10 (West 1996 & Supp. 2000, 2001). Although these statutes are much broader than other states' fetal homicide laws, they have withstood constitutional challenge in state court. *See, e.g.*, *Merrill*, 450 N.W.2d 318; *State v. Bauer*, 471 N.W.2d 363 (Minn. Ct. App. 1991).

[32] *Compare, e.g.*, IOWA CODE ANN. § 707.7 (1993 & Supp. 1996) (criminalizing consensual feticide "after the end of the second trimester of the pregnancy") *with id.* § 707.8 (making the termination of a "human pregnancy without the consent of the pregnant person" a felony without specifying gestational limits).

[33] *See Keeler*, 470 P.2d at 624.

[34] *See* Julie N. Qureshi, People v. Davis*: California's Murder Statute and the Requirement of Viability for Fetal Murder*, 25 GOLDEN GATE U.L. REV. 579, 580 (1995).

[35] *See* CAL. PENAL CODE § 187(a) (West 1999).

[36] *See Horne*, 319 S.E.2d at 704.

[37] S.C. CODE ANN. § 16-3-10 (Law. Co-op. 1976 & West Supp. 1998).

[38] Colleen Jolicoeur-Wonnacott, *The Unborn Victims of Violence Act: Friend or Foe to the Unborn?*, 17 T. M. COOLEY L. REV. 563, 576 (2000) (discussing the Unborn Victims of Violence Act of 1999); *see also* H.R. REP. No. 107-42, pt. 1 at 78.

[39] *See, e.g.*, GA CODE ANN. § 16-5-80(a) (1982).

[40] *See* H.R. 503, §§ 2–3.

intent,"[41] a common law doctrine that dictates, "a defendant[,] who intends to kill one person but instead kills a bystander, is deemed the author of whatever kind of homicide would have been committed had he killed the intended victim."[42] Thereby, under the UVV, when an assailant commits an unlawful act against a pregnant woman and, in so doing, harms the fetus, the unlawful intent toward the mother is "transferred" to the fetus. The assailant is then punished using the federal law that would apply if "the injury or death occurred to the unborn child's mother."[43] If, however, the harm or attempted harm to the fetus is intentional, the assailant is punished according to the Federal Criminal Code provisions for "intentionally killing or attempting to kill a human being."[44] Thus, the UVV treats fetal life as human life—that is, the law treats an offense against a fetus as it would an offense against the mother—when prosecuting crimes committed against pregnant women.[45]

Although the UVV classifies fetuses as human beings for the purposes of third party criminal acts directed against pregnant women, the Act explicitly declines to criminalize acts of certain specified parties. The Act does not apply to "any person for conduct relating to an abortion for which the consent of the pregnant woman" or any authorized party has been obtained.[46] The Act further provides that those providing medical treatment to the woman or her unborn child will not be subject to the UVV's penalties.[47] Finally, the acts of "any woman with respect to her unborn child" also do not fall under the Act.[48] The UVV, therefore, facially excuses consensual abortions.[49]

---

[41] *Id.*

[42] Jolicoeur-Wonnacott, *supra* note 38, at 567 (quoting H.R. REP. No. 106-332, pt. 1, at 10 (106th Cong., 1999)).

[43] H.R. 503, §§ 2–3.

[44] *Id.* The Federal Criminal Code provisions referred to in the UVV are 18 U.S.C. §§ 1111–1113 (1994 & Supp. III 1998). *Id.*

[45] *See* H.R. 503, §§ 2–3. As mentioned earlier, the Act, however, explicitly proscribes a death sentence for fetal homicide. *See id.*

[46] *Id.*

[47] *See id.*

[48] *Id.*

[49] The Act thus implicitly prohibits prosecuting the mother for such potentially harmful conduct as drug or alcohol abuse during pregnancy. *Compare* Lynn M. Paltrow, *Pregnant Drug Users, Fetal Persons, and the Threat to* Roe v. Wade, 62 ALBANY L. REV. 999 (1999) (stating that most states do not impose criminal punishments on women for their actions during pregnanxy but that some state courts have recently begun to take this drastic step), *and* Alma Tolliver, *Child Abuse Statute Expanded to Protect the Viable Fetus: The Abusive Effects of South Carolina's Interpretation of the Word "Child,"* 24 S. ILL. U.L.J. 383 (2000) (noting that most states do not "extend third person liability to the mother of the viable fetus" for such harmful actions as drug abuse during pregnancy), *with Whitner v. State*, 492 S.E.2d 777 (S.C. 1997) (holding a mother criminally liable for using illegal drugs while pregnant under a child abuse endangerment statute, which, according to the court, included "fetus" in its definition of "child").

Identifying egregious examples of criminal acts against pregnant women,[50] supporters of the UVV contend that it ensures that "criminals who commit violent acts against pregnant women are justly punished for injuring or killing unborn children."[51] This viewpoint is often emphasized in the context of domestic violence, as many of the leading cases on feticide involve spousal abuse.[52] Proponents of the bill believe that increasing penalties for feticide will deter domestic battery.[53] Further, some argue that fetal homicide statutes will allow for detection and punishment of spousal battering in instances where a pregnant woman would otherwise fail to report abuse.[54] The Act will provide prosecutors with "a tool to reach and punish the partner without having to depend upon the mother's cooperation" in pressing charges.[55]

Critics of the Act, however, assert that it does not adequately protect battered women.[56] During congressional hearings on the issue, Juley Fulcher of the National Coalition Against Domestic Violence ("NCADV") argued, first, that the UVV would shift the criminal law's focus from violence against women to crimes against the unborn fetus.[57] Second, Fulcher claimed that the Act would be underprotective of domestic violence victims since the law only applies to pregnant women in federal cases.[58] The UVV would also do little, if anything, to address the overall prosecution of domestic violence: the majority of domestic violence crimes are committed against non-pregnant women at the state level.[59] Finally, Fulcher reasoned, a battered woman may be even more intimidated by her batterer into hiding her miscarriage and foregoing medical treatment if taking these actions could result in her batterer's prosecution

---

[50] *See Legislative Hearing on H.R. 503 the "Unborn Victims of Violence Act of 2001" Before the Subcomm. on the Constitution of the House Comm. on the Judiciary*, 107th Cong. 13 (Mar. 15, 2001) [hereinafter *Hearing on the UVV*] (statement of Rep. Steve Chabot, Chairman). Chabot uses examples such as the case of Reginald Anthony Falice, who shot and killed his pregnant wife and her unborn child as she sat at a red light. *See id.* Falice was convicted under federal law of interstate domestic violence and using a firearm while committing a violent crime. *See id.* He was not convicted or punished for killing his unborn fetus, however, because federal law did not recognize fetuses as victims of murder. *See id.*; *see also United States v. Falice*, No. 00-4559, 2001 WL 1082447 (4th Cir. Sept. 17, 2001).

[51] *Hearing on the UVV*, *supra* note 50, at 14–15 (statement of Rep. Steve Chabot).

[52] *See* Michael S. Robbins, *The Fetal Protection Act: Redefining "Person" for the Purposes of Arkansas' Criminal Homicide Statutes*, 54 ARK. L. REV. 75, 87 (2001). Although the statute is, in part, justified as a means of protecting battered women, the UVV should not have any trouble passing constitutional scrutiny under *United States v. Morrison*, 529 U.S. 598 (2000), since "the bill does not extend Congress reach to prohibit any conduct that does not currently violate Federal law . . . [and] thus relies on the predicate crimes for its constitutional hook." H.R. REP. NO. 107-42, pt. 1, at 10–11.

[53] *See* Robbins, *supra* note 52, at 87.

[54] *See id.*

[55] *Id.*

[56] *See Hearing on the UVV*, *supra* note 50, at 52 (prepared statement of Juley Fulcher, Public Policy Dir., National Coalition Against Domestic Violence).

[57] *See id.*

[58] *See id.*

[59] *See id.*

for murder.[60] Therefore, "[t]he long-term public health implications of such a policy would," according to Fulcher, "be devastating for victims of domestic violence and all women."[61]

Despite these divisive policy debates, the most contentious issue surrounding the UVV is how it will affect the right to abortion as defined nearly thirty years ago by *Roe v. Wade*. The Court held in *Roe* that "the word 'person,' as used in the Fourteenth Amendment, does not include the unborn."[62] By contrast, the UVV specifically defines a fetus as "a member of the species homo sapiens" and as a "child in utero."[63] These phrases and the penalties applied by the Act equate a fetus with a human being and a child, differentiating the entities only by placement of the fetus within its mother's womb. If a fetus is a "person" under criminal statutes, critics of fetal homicide laws argue, then "presumably fetuses have a right to life under the Fourteenth Amendment."[64] This logical extension of the UVV directly contradicts the foundations upon which *Roe* allows consensual abortions.

The United States Supreme Court's 1973 decision in *Roe v. Wade* established a woman's right to choose to abort a nonviable fetus.[65] In *Roe*, the Court considered Texas's argument that a fetus was a "person" under the Due Process Clause of the Fourteenth Amendment.[66] The Court gave three reasons for holding that a fetus is not such a "person." First, there was no case law that supported a finding of fetal personhood under the Fourteenth Amendment.[67] Second, none of the uses of "person" in the Constitution had any application before birth.[68] Finally, the history of abortion laws showed that they granted a much broader right to abortion in the nineteenth century, and thus in 1868 when the Fourteenth Amendment was adopted, than they did in the 1970s when the Court decided *Roe*; this suggested that the Amendment's drafters did not intend to include fetuses in its protections.[69] The Court also found persuasive the fact that declaring a fetus a protected "person" under the Fourteenth Amendment would contradict existing abortion laws that provided exceptions for the life and health of the mother; the value of a fetus's life could not be equal to that of a full person if a fetus could legally be killed in cer-

---

[60] *See id.*

[61] *Id.*

[62] *Roe*, 410 U.S at 158. The Fourteenth Amendment states that "All persons *born* or naturalized in the United States, and subject to the jurisdiction thereof are citizens of the United States," and further that "No State shall . . . deprive any person of life, liberty, or property, without due process of law." U.S. CONST. amend. XIV, § 1 (emphasis added).

[63] H.R. 503, §§ 2–3.

[64] Tsao, *supra* note 22, at 470.

[65] *See Roe*, 410 U.S. at 163–65.

[66] *See id.* at 156–58.

[67] *See id.* at 157.

[68] *See id.*

[69] *See id.* at 158; *see also* Agota Peterfy, *Fetal Viability as a Threshold to Personhood*, 16 J. LEGAL MED. 607, 611 (1995).

tain instances.[70] The Court declined to adopt any theory of when life began,[71] but it did recognize that the law, outside of abortion, was "reluctant to endorse any theory that life . . . begins before life [sic] birth or to accord legal rights to the unborn."[72]

Given that a fetus was not a full person, the *Roe* Court held that the right to privacy previously established in constitutional jurisprudence[73] encompassed a woman's decision to terminate her pregnancy.[74] Writing for the Court, Justice Blackmun used a trimester framework to balance the woman's privacy right against the state's interests in protecting the health and safety of the pregnant woman and the potential life of the fetus.[75] When those state interests became compelling, they could support narrowly tailored regulations of abortion that furthered purposes related to them.[76]

Blackmun's trimester framework sought to balance state interests and personal privacy rights as a fetus advanced in gestational age. In the first trimester, when abortion may be safer than carrying a fetus to term, a woman could abort her fetus free from state interference or regulation.[77] During this time, the state's interests were not "compelling," so the state could not justify infringing on the right to abortion.[78] In the second trimester, however, as abortions became medically more dangerous, the

---

[70] *See* Roe, 410 U.S. at 157 n.54; *see also* Ronald Dworkin, *Unenumerated Rights: Whether and How* Roe *Should Be Overruled*, 59 U. CHI. L. REV. 381, 398–99 (explaining that "[i]f a fetus is a constitutional person, then states not only may forbid abortion but, at least in some circumstances, must do so. No justice . . . has even advanced that claim."). One might consider footnote fifty-four to be the *Roe* Court's indication that fetal status and fetal rights laws should be made more consistent. Precisely, it begs the question of how the law can protect a fetus's rights so fiercely in some contexts but not in others.

[71] *See Roe*, 410 U.S. at 159.

[72] *Id.* at 161.

[73] For development of the right of privacy, see, for example, Griswold v. Connecticut, 381 U.S. 479 (1965) (holding that the right of privacy found in the Bill of Rights protects married couples against a state's prohibition on contraceptive use); Eisenstadt v. Baird, 405 U.S. 438 (1972) (holding that the Equal Protection Clause extends the privacy right to unmarried people as well). Although not specifically mentioned in the text of the Constitution, federal jurisprudence holds that the right to privacy is part of the Liberty protected by the Due Process Clause and the "penumbras" around sections of the Bill of Rights. *Griswold*, 381 U.S. at 484–86 (defining "penumbras" as "emanations from [Bill of Rights] guarantees that help give them life and substance."). The privacy right extends to marriage, procreation, contraceptive use, and family relationships, among other protected areas. *See* Carey v. Population Serv. Int'l, 431 U.S. 678, 684–85 (1977).

[74] *See Roe*, 410 U.S. at 153–55.

[75] *See id.* at 163–65.

[76] *See id.* Fourteenth Amendment jurisprudence holds the right of privacy, including the right to choose abortion, to be fundamental. *See* Rachael K. Pirner & Laurie B. Williams, Roe *to* Casey: *A Survey of Abortion Law*, 32 WASHBURN L.J. 166, 168. Thus, it falls under the prescriptions that "where certain 'fundamental rights' are involved . . . regulation limiting these rights may be justified only by a 'compelling state interest,' and . . . legislative enactments must be narrowly drawn to express only the legitimate state interests at stake." *Roe*, 410 U.S. at 155 (citations omitted). As applied to *Roe*, Blackmun indicated that the two-prong test guarantees the right to abortion absent a compelling state interest and narrowly tailored regulation. *See id.* at 156.

[77] *See Roe*, 410 U.S. at 163.

[78] *Id.* (internal quotations omitted).

state's interest in protecting the health and safety of the mother became compelling.[79] The state could then regulate abortion in ways reasonably related to "the preservation and protection of maternal health"; any regulations justified by the compelling interest of protecting a mother's health, however, had to be narrowly tailored to further only that interest.[80] Following viability,[81] occurring at approximately the end of the second trimester, a state could prohibit abortion as long as it made exceptions for pregnancies that could endanger the life or health of the mother.[82] Because only a viable fetus could maintain an independent existence outside of the womb, the state's interest in protecting potential life could not justify prohibitions on abortion until after viability.[83]

Successor cases to *Roe* have modified and defined the contours of the right to choose abortion, but they have refused to overturn *Roe* explicitly.[84] The first plurality opinion to challenge *Roe* directly was in *Webster v. Reproductive Health Services*.[85] *Webster* held that states could adopt their own theories of when life began, so long as those theories were not used to justify restrictions on abortion that would otherwise be invalid under *Roe*.[86] The Court thus declined to declare unconstitutional a Missouri statute whose preamble explicitly stated that life began at conception.[87]

Furthermore, in contrast to *Roe*'s holding that a state could not justify interfering with the abortion of a pre-viable fetus based on the state's interest in protecting prenatal life, *Webster* permitted a state to promote that interest before viability as long as the regulation had the purpose of protecting a fetus that might have been viable.[88] The statute at issue in *Webster* mandated that a physician conduct certain viability tests on a fetus before performing an abortion for a woman whom the physician believed was more than twenty weeks pregnant.[89] The District Court, however, specifically found that "the medical evidence is uncontradicted

---

[79] *See id.*

[80] *Id.*

[81] Justice Blackmun defined viability as the point at which a fetus is "potentially able to live outside the mother's womb, albeit with artificial aid." *Id.* at 160. In 1973, viability usually fell between the twenty-fourth and twenty-eighth week of gestation. *See id.*

[82] *See id.* at 163–64.

[83] *See id.*

[84] Three of the current Justices—Rehnquist, Scalia, and Thomas—have indicated that *Roe* should be overturned. *See* Planned Parenthood of Southeastern Pennsylvania v. Casey, 505 U.S. 833, 944 (1992) (Rehnquist, C.J., dissenting). Rehnquist's concurrence was also joined by Justice White who has since retired from the Court. *See id.*

[85] 492 U.S. 490 (1989). Although the Court refused to overturn *Roe* in *Webster*, Rehnquist's plurality opinion in that case sounded a clear call for the Court to do so at its next opportunity. *See id.* at 518–20. Justice Scalia's concurrence carried a similar exhortation. *See id.* at 532 (Scalia, J., dissenting).

[86] *See id.* at 506.

[87] *See id.* at 504, 507.

[88] *See id.* at 519–20.

[89] *See id.* at 501.

that a 20-week fetus is *not* viable," thus indicating that the regulation at issue actually required interference with abortions before viability and regardless of any concern for promoting the life or health of the mother.[90] The *Webster* Court still upheld the regulation, stating that there was no reason "why the State's interest in protecting potential human life should come into existence only at the point of viability." Thereby, the Court invalidated a major portion of the *Roe* framework, while refusing to overturn it explicitly.[91]

In *Planned Parenthood of Southeastern Pennsylvania v. Casey*,[92] the Court adopted a new standard for abortion law, while simultaneously reaffirming what the plurality authors considered the "essential holding" of *Roe*.[93] The *Casey* Court reiterated that a woman can abort her nonviable fetus without excessive state interference, but that states can choose to prohibit abortions of viable fetuses as long as they provide exceptions for the life and health of the mother.[94] The Court further acknowledged that states may have legitimate interests in protecting the health of the woman and the potential life of the fetus throughout pregnancy.[95] While affirming those elements of *Roe*, however, the plurality abandoned *Roe*'s trimester framework, finding that it misinterpreted the nature of the pregnant woman's interests and undervalued the state's interest in potential life.[96] In its place, *Casey* instituted the "undue burden" standard, which

---

[90] *Id.* at 515 (internal quotations omitted).

[91] *Id.* at 519, 521. The plurality opinion did, however, state that *Roe* was "unsound in principle and unworkable in practice," calling for a case with appropriate facts on which to overturn the precedent. *Id.* at 494–95 (quoting Garcia v. San Antonio Metropolitan Transit Authority, 469 U.S. 528, 546 (1985)). The plurality considered *Roe* unsound and unworkable because its rigidity was inconsistent with a Constitution cast in general terms and because, having held that a state could assert its interest in potential life throughout pregnancy, *Roe*'s focus on viability as a dividing line was no longer relevant. *See id.* Because the facts of *Webster* differed too much from those of *Roe*, however, the plurality refused to overturn *Roe* at that time. *See id.* at 495. *But see id.* at 538 (Blackmun, J., dissenting) (stating that "[t]he simple truth is that *Roe* would not survive the plurality's analysis").

[92] In *Casey*, plaintiffs challenged the constitutionality of the 1988 and 1989 amendments to the Pennsylvania abortion statute. The amendments contained the following requirements: a woman seeking an abortion must give her informed consent to the procedure; a woman must receive specified information at least twenty-four hours before the abortion procedure; a minor seeking an abortion must have either the consent of one of her parents or a judicial bypass; a married woman seeking an abortion must provide a signed statement that she told her husband about the abortion; and abortion facilities must comply with certain reporting requirements. *See Casey*, 505 U.S. at 844.

[93] *Id.* at 846.

[94] *See id.*

[95] *See id.* Justice O'Connor, writing for the plurality, emphasized that these "principles do not contradict one another." *Id.* Even though the state has legitimate interests in the fetus's potential life and the mother's health throughout pregnancy, before viability, those interests "are not strong enough to support a prohibition of abortion or the imposition of a substantial obstacle to the woman's effective right to elect the procedure." *Id.*

[96] *See Casey*, 505 U.S. at 873. Critics of *Casey* could argue that the "essential holding" of *Roe* included the abandoned trimester framework. *See, e.g., id.* at 994 (Scalia, J., concurring in part and dissenting in part) ("I must confess, however, that I have always thought, and I think a lot of other people have always thought, that the arbitrary trimester

presumed that a state's regulation of abortion was constitutional throughout pregnancy unless it had "the purpose or effect of placing a substantial obstacle in the path of a woman seeking an abortion of a nonviable fetus."[97]

Despite the fact the Supreme Court has restricted the breadth of abortion rights since *Roe*, none of its decisions has contested the finding that a fetus is not a full person under the law or the Constitution.[98] According to Justice Stevens,

> [N]o Member of the Court has ever questioned this fundamental proposition. Thus, as a matter of federal constitutional law, a developing organism that is not yet a "person" does not have what is sometimes described as a "right to life." This . . . remains a fundamental premise of our constitutional law governing reproductive autonomy.[99]

The UVV would be the first federal law, in any legal context, to question this underlying proposition. Though supporters maintain that

---

framework, which the Court today discards, was quite as central to *Roe* as the arbitrary viability test, which the Court today retains.").

[97] *Id.* at 877 (additionally stating that a "statute with this purpose is invalid because the means chosen by the State to further the interest in potential life must be calculated to inform the woman's free choice, not hinder it"). Justice O'Connor noted, however, that a statute that "has the incidental effect of making it more difficult or more expensive" to have an abortion is not invalid; only a statute posing an "undue burden" would be unconstitutional. *Id.* at 874. In this way, statutes that piled increasing numbers of regulations onto the abortion procedure could be seen as merely making it more "difficult" or "expensive" to exercise the right to choose, while not reaching into the "heart of the liberty protected by the Due Process Clause." *Id. Cf.* John L. Horan, *A Jurisprudence of Doubt:* Planned Parenthood v. Casey, 26 CREIGHTON L. REV. 479, 519 (1993) (noting a shift in burden of proof from the state to the party challenging a regulation's constitutionality).

[98] *See* Pirner & Williams, *supra* note 76, at 166, 171 (noting that "no Supreme Court Justice has ever adopted [the] position" that "a fetus is a person within the meaning of the Fourteenth Amendment").

[99] *Casey*, 505 U.S. at 913–14 (Stevens, J., concurring in part and dissenting in part); *see also Webster*, 492 U.S. at 569 n.13 (Stevens, J., dissenting in part and concurring in part). Even though the *Webster* plurality refused to declare unconstitutional a statute that stated that life began at conception, *see id.* at 507, this refusal is inherently different from declaring a fetus to be a "person" under the Fourteenth Amendment, and thus does not alter *Roe*'s holding. Ronald Dworkin addresses the idea that individual states could declare a fetus to be a "person," even if constitutional jurisprudence did not:

> That suggestion assumes that a state can curtail some persons' constitutional rights by adding new persons to the constitutional population. The constitutional rights of one citizen are of course very much affected by who or what else also has constitutional rights, because the rights of others may compete or conflict with his. So any power to increase the constitutional population by unilateral decision would be, in effect, a power to decrease rights the national Constitution grants to others . . . . [I]f a state could declare a fetus a constitutional person, it could prohibit abortion even when the pregnancy threatens the mother's life, just as it normally forbids killing one innocent person to save the life of another.

Dworkin, *supra* note 70, at 400–02.

the UVV "in no way interferes with or restricts the abortion right articulated in *Roe*,"[100] critics of the Act contend that a fetus's legal status should not be context dependant. They argue that a fetus should be treated the same in other legal arenas as it is in abortion law.[101] To consider a fetus to be a person for the purposes of criminal law—as the UVV does—but not abortion law, would be to create a mere legal fiction.

Congressional proponents of the UVV argue that, although it did not consider the fetus a "person[ ] in the whole sense," *Roe* "recogniz[ed] unborn children as persons for purposes other than abortion, such as inheritance and tort injury, purposes which the *Roe* court itself recognized as legitimate."[102] This contention, however, misinterprets *Roe*'s discussion of these limited fetal rights. *Roe* specifically noted that legal rights have been accorded to the unborn only "in narrowly defined situations" or where "the rights are contingent upon life [sic] birth."[103] For example, wrongful death suits may be brought on behalf of stillborn children, but, according to the *Roe* Court, "such an action . . . would appear . . . to vindicate the parents' interest and is thus consistent with the view that the fetus, at most, represents only the potentiality of life."[104] Similarly, the Court noted that some states recognize unborn children's property or inheritance rights, but that the "perfection of the interests involved . . . has generally been contingent upon live birth."[105] Such "rights," according to *Roe*, did not alter the proposition that a fetus is not a constitutional "person."[106] Instead, the fact that a fetus receives only some very narrow rights reinforced the idea that a fetus is not a full "person." In other words, a fetus may be accorded some sticks in the bundle of rights that constitute personhood, but it never acquires the full bundle. As a result, a fetus should not be considered a full "person" in a broad area of the law like criminal law—as the UVV would have it—so long as other areas of the law continue to consider a fetus as less than a person.[107]

In *Webster*, however, the plurality relied on *Roe*'s recognition of limited fetal tort and probate rights to hold that, so long as there is no interference with the right to abortion, a state may adopt its own theory

---

[100] H.R. REP. No. 107-42, pt. 1, at 12.

[101] *Cf.* Peterfy, *supra* note 69, at 611 (claiming that other courts have used *Roe*'s reasoning to exclude fetuses from wrongful death and homicide statutes).

[102] H.R. REP. No. 107-42, pt. 1, at 12. *See also Roe*, 410 U.S. at 162 (noting that "unborn children have been recognized as acquiring rights or interests by way of inheritance or other devolution of property").

[103] Roe, 410 U.S. at 161.

[104] *Id.* at 162.

[105] *Id.*

[106] *See id.*

[107] This statement does not deny the difference between abortion and fetal murder, in that abortion poses a fetus's "rights" against the mother's while homicide does not. The changed circumstances are not the crucial point. Instead, the point is that the underlying law has not changed, so that the definition of "person" under the Constitution should remain constant regardless of the surrounding situation.

of when life begins.[108] Rather than viewing *Roe*'s discussion of tort and probate rights as limiting fetal rights, the *Webster* Court used these rights as examples of the state's expansive reach in defining human life, as long as state definitions did not directly conflict with the right to abortion.[109] While *Webster*'s interpretation fundamentally misconstrued *Roe*, the Supreme Court could use *Webster* to uphold the UVV as constitutional.[110] Giving the states such latitude in defining "human life" suggests that the federal government could espouse similarly broad definitions in its criminal law.

Furthermore, supporters of the UVV and similar state fetal homicide laws claim that, rather than threatening the right to choose an abortion, the Act serves instead to support that choice. By specifically exempting abortion from prohibited conduct, the Act seems to explicitly confirm the abortion right.[111] Moreover, once a woman makes the fundamental choice, the UVV requires the federal government to give heightened protection to a woman's decision to carry her pregnancy to term.[112] Justice Kennard's concurrence in *People v. Davis* reflects this idea:

> [W]hen a fetus dies as the result of a criminal assault on a pregnant woman, the state's interest extends beyond the protection of potential human life. The state has an interest in punishing violent conduct that deprives a pregnant woman of her procreative choice.[113]

Under this reasoning, the Act actually reaffirms *Roe*, or at least leaves the right to abortion unscathed.

Many argue, however, that this rationale is "not entirely satisfactory."[114] First, the protection of the fetus during pregnancy is entirely disconnected with the woman's "choice" to continue her pregnancy.[115] Her

---

[108] *See Webster*, 492 U.S. at 506.

[109] *See id.*

[110] *See* H.R. Rep. No 107-42, pt. 1, at 12. This was the rationale employed in *Merrill*, which upheld a conviction for fetal homicide, despite the fact that *Roe* determined that an unborn child lacks personhood. *See Merrill*, 450 N.W.2d at 322. The Minnesota Supreme Court maintained that "[t]he right in *Roe v. Wade* can be understood only by considering both the woman's interest and the nature of the State's interference with it," and further noted that, under *Webster*, "[e]ven laws which directly impact on abortion are constitutional so long as the statute itself does not impinge on the woman's decision." *Id.* As a result, the *Merrill* court held that, so long as a woman's privacy rights are not directly abridged, fetal homicide statutes are constitutional. *See id.*

[111] *See* Robbins, *supra* note 52, at 87–88.

[112] *See id.* ("Pro-choice advocates want women to continue to have the right to abort an unwanted pregnancy but are opposed to forced abortions, hence the label 'choice.' . . . All women, even pro-choice advocates, who choose to bear a child, as opposed to aborting the pregnancy, want that unborn child to be protected from harm from third-party attackers.").

[113] *Davis*, 872 P.2d at 604 (Kennard, J., concurring).

[114] *See* Tsao, *supra* note 22, at 470.

[115] *See id.*

choice has already been made, has already been protected by law, and, if her fetus is already viable, may be unalterable anyway. The UVV, instead, protects the physical integrity of a fetus, regardless of the mother's choice as to whether to have an abortion.[116] Furthermore, the UVV explicitly protects the life of the fetus, not the woman's procreative choice. A focus on the rights of the fetus, rather than on the right of a pregnant woman to be free from bodily injury, also weakens the framework on which abortion is premised and may, therefore, further compromise the right to abortion.

In addition to its inconsistencies with abortion jurisprudence, the UVV will also likely face several other constitutional challenges. Recent legal attacks on similar state statutes suggest that the UVV may be vulnerable on Fifth Amendment equal protection and due process grounds,[117] as well as to claims based on the Eighth Amendment's[118] prohibition of cruel and unusual punishment.[119] While these challenges may have some merit, state courts—the only courts so far to have addressed feticide statutes—have uniformly rejected them, maintaining that, so long as feticide laws do not criminalize protected abortions, states have considerable latitude in shaping and defining regulations to protect the unborn. Although state decisions are by no means dispositive, that state courts have consistently upheld feticide statutes may be persuasive if the UVV is challenged in federal court. The UVV is, on the other hand, conceived more broadly than many state laws—by allowing punishment at any stage of gestation regardless of criminal intent—and may therefore be constitutionally problematic.

Because the UVV immunizes mothers who harm their unborn fetuses from criminal prosecution but makes third party attackers liable for similar actions, the UVV may violate the Fifth Amendment's equal protection guarantee.[120] To establish a violation of equal protection of the law, a claimant must show that he or she is being treated differently from someone else who is similarly situated.[121] Thus, under the UVV, it is possible that, "if a mother shoots herself in the stomach with the intention of killing her unborn child, she would not be prosecuted," whereas a third party attacker would be punished for the same act.[122] The mother's special status and right to privacy have been recognized by the Supreme

---

[116] *See id.*

[117] *See* U.S. CONST. amend. V. State cases have addresses statutes' constitutionality under the Fourteenth Amendment, which applies to states. The older Fifth Amendment imposes the same mandates on the federal government. Since the UVV is a federal law, the Fifth Amendment would apply. *See* U.S. CONST. amend. V. Constitutional jurisprudence is substantially the same in its analysis of the two amendments.

[118] U.S. CONST. amend. VIII.

[119] *See* Robbins, *supra* note 52, at 88–89.

[120] U.S. CONST. amend. XIV, § 1.

[121] *See Parham v. Hughes*, 441 U.S. 347, 354–55 (1979).

[122] Robbins, *supra* note 52, at 89.

Court,[123] however, so the argument that a third party assailant should be treated similarly to a pregnant woman with respect to her unborn fetus is unlikely to invalidate the Act.

State courts have, in fact, overwhelmingly rejected this claim, determining that third party attackers are not similarly situated to pregnant women.[124] In *State v. Merrill*, a leading case on the constitutionality of feticide statutes, the Minnesota Supreme Court rejected an equal protection claim made by a defendant who shot and killed a pregnant woman, allegedly causing the death of a twenty-seven or twenty-eight-day-old embryo.[125] The defendant claimed that the state violated the Equal Protection Clause by subjecting him to prosecution for terminating a pregnancy while allowing others—namely, pregnant women in the abortion context—to terminate fetal life without criminal consequences.[126] In comparing his rights to those of pregnant women, the defendant claimed that "similarly situated persons [were] treated dissimilarly" under the law.[127] The court rejected his argument, reasoning that *Roe* protects only a woman's constitutional right to decide whether to terminate her own pregnancy, not another person's right to terminate that pregnancy.[128] By virtue of this unique maternal right, a pregnant woman is not similarly situated to a third party actor.[129]

A version of the equal protection argument was also addressed by the California Supreme Court in *People v. Davis*.[130] In *Davis*, the defendant shot a pregnant woman in the chest when she resisted an attempted robbery.[131] The woman, who was between twenty-three and twenty-five weeks pregnant, miscarried as a result of the gunshot wound and consequent loss of blood.[132] The defendant was convicted of felony murder for the death of the fetus during the commission of theft,[133] and appealed, claiming that the jury must find that the fetus was viable before it could find the defendant guilty under California's murder statute.[134] The defen-

---

[123] *See Roe*, 410 U.S. at 153.

[124] *See, e.g., Merrill*, 450 N.W.2d 318; *Davis*, 872 P.2d 591.

[125] *See Merrill*, 450 N.W.2d at 318.

[126] *See id.* The defendant also argued that the state violated Equal Protection by "adopt[ing] a classification equating viable fetuses and nonviable embryos with a person." *Id.* Because *Roe* says that an unborn child is not a "person" under the Fourteenth Amendment, the defendant contended that the Minnesota criminal statute violated the Equal Protection Clause by treating the two distinct entities as equivalent. *See id.* The defendant did not, however, have standing to make this claim, so the court did not address his argument. *See id.*

[127] *Id.*

[128] *See id.* at 321–22.

[129] *See id.* at 321–22.

[130] *See Davis*, 872 P.2d at 597.

[131] *See id.* at 593.

[132] *See id.*

[133] *See id.*

[134] *See id.* at 594. California's murder statute provides that: "Murder is the unlawful killing of a human being, or a fetus, with malice aforethought." CAL. PENAL CODE § 187(a)

dant, like the defendant in *Merrill*, claimed that under *Roe*, the victim's fetus could have been legally aborted, and therefore, that he could not be prosecuted for its murder.[135] The California Supreme Court, however, held that "a defendant who intentionally murders a fetus, and a pregnant woman who chooses to terminate her pregnancy, are not similarly situated" since "[a] woman has a privacy interest in terminating her pregnancy [whereas a] defendant has no such interest."[136] Because, under *Roe*, the state still had an interest in protecting the potential life of the fetus prior to viability, the defendant could be criminally charged for fetal murder at any point after the post-embryo period.[137]

Although the UVV will likely withstand equal protection scrutiny, it may additionally face attack under the Fifth's Due Process Clause[138] on vagueness grounds. Procedural due process mandates that criminal laws give people sufficient notice of prohibited behavior in order to protect them from arbitrary and capricious enforcement of the law: individuals must be warned of "prohibited actions so that they may conduct themselves accordingly."[139] The UVV does not require that the fetus be viable, nor does it require a defendant to know of the victim's pregnancy.[140] Critics of the UVV contend its criminalization of fetal homicide—regardless of a defendant's intent or awareness of the risk that feticide may occur—deprives a defendant of any notice that commission of certain actions may make him or her liable for murder.[141]

Justice Mosk's dissent in *Davis* most convincingly presents this point of view.[142] Justice Mosk contended that to hold a defendant criminally liable for the unintended death of a "nonviable and invisible fetus" would violate his or her due process rights.[143] Justice Mosk emphasized that the fetus was "literally invisible to everyone," meaning that the de-

---

(West 1999). Before *Davis*, all of the California courts of appeals that had addressed the issue had held that viability was a prerequisite for fetal homicide. *See* Folger, *supra* note 16, at 238; *see also* People v. Smith, 129 Cal. Rptr. 498, 502 (Cal. Ct. App. 1976); People v. Apodaca, 142 Cal. Rptr. 830, 836 (Cal. Ct. App. 1978); People v. Henderson, 275 Cal. Rptr. 837, 853–54 (Cal. Ct. App. 1990).

[135] *See Davis*, 872 P.2d at 597.

[136] *Id.* at 598 (quoting *People v. Ford*, 581 N.E.2d 1189, 1199 (Ill. App. Ct. 1991)) (internal quotations omitted).

[137] *See id.* at 597. The defendant in *Davis*, however, was not convicted since the rule announced in the case was a change from the previous California viability rule, and this defendant's conviction would therefore be a violation of his Due Process rights and his rights under the Ex Post Facto Clause of the Constitution, U.S. CONST. Art. I, § 9. *See Davis*, 872 P.2d at 600.

[138] U.S. CONST. amend. V. The federal Due Process clause reads: "No person shall be . . . deprived of life, liberty, or property, without due process of law." *Id.*

[139] Tsao, *supra* note 22, at 472. *See also* Kolander v. Lawson, 461 U.S. 352, 357 (1983).

[140] *See* H.R. 503, §§ 2–3.

[141] *See* H.R. Rep. No. 107-42, pt. 1, at 81.

[142] *Davis*, 872 P.2d at 607 (Mosk, J., dissenting).

[143] *Id.* at 619.

fendant could not even know that the fetus existed.[144] Moreover, he noted, if the victim is "one of many women with some irregularity in her menstrual cycle, she herself may not know she is pregnant."[145] Without knowledge of an existing pregnancy, the defendant would not have sufficient notice of the crime for which he could be prosecuted. Finally, Justice Mosk pointed out the particular perils of imposing such a vague feticide law on top of the felony murder rule, which already eliminates the *mens rea* requirement for murders committed in the course of a felony.[146] He concluded that extending the felony murder rule to fetal killing would lead to "draconian" punishment of defendants, regardless of knowledge or intent.[147] Without a viability requirement, the UVV may be too vague to provide adequate notice to a defendant of his crime because it would allow attackers to be prosecuted for murder even during the very early stages of a woman's pregnancy, when the woman's condition may not be apparent to third parties or even to herself.

Moreover, feticide statutes may be unconstitutionally vague because of the inherent difficulties of proving causation in feticide cases.[148] Justice Mosk noted "spontaneous abortions,"[149] often caused by a variety of health factors in the mother and the fetus, occur in fifteen to twenty percent of all pregnancies.[150] The *Davis* majority, according to Justice Mosk,

---

[144] *Id.* at 615. The pregnant woman, Maria Flores, was five feet and one inch tall and weighed 191 pounds. *Id.* Based on these measurements, the perinatologist testified at trial that it was "'not likely' that on the date of the shooting a woman of Flores's stature would have showed her pregnancy when clothed and standing upright." *Id.* Justice Mosk concluded: "I cannot believe that the Legislature intended to make it murder—indeed, capital murder, to cause the death of an object the size of a peanut." *Id.*

[145] *Id.*

[146] *See id.* at 619.

[147] *Id.* Justice Mosk offered a hypothetical to illustrate the dangerous implications of the majority opinion:

> [A]n unarmed 18-year-old with no criminal record enters a store during business hours, intending to shoplift a can of spray paint; when a security guard accosts him, his nerve fails and he bolts for the door; in his haste he accidentally knocks a woman shopper to the floor; unknown to anyone, the woman is 7 weeks' pregnant, and the trauma of the fall causes her to miscarry. Before today's decision, such a youth would be guilty at most of second degree burglary . . . and would be punishable by either county jail time of up to one year or a prison term of sixteen months or two or three years. . . . After today's decision, however, this teenager could also be found guilty of first degree murder of the fetus on a burglary felony-murder theory; . . . in that event his punishment would at least be imprisonment for 25 years to life.

Id. (citations omitted).

[148] *See id.* at 620; *see also* Mark S. Kende, *Michigan's Proposed Prenatal Act: Undermining A Woman's Right to an Abortion*, 5 AM. U.J. GENDER & L. 247, 257–58 (1996).

[149] *Davis*, 872 P.2d at 620 (Mosk, J., dissenting). "Spontaneous abortion is the medical term for miscarriage." *Id.* at 620 n.14.

[150] *See Davis*, 872 P.2d at 620 (including "genetic or developmental defects in the fetus, uterine abnormalities, maternal trauma, illness, or substance abuse, [and] toxins in the fetal or maternal environment"). Some commentators estimate that up to "seventy-eight

overlooked the "medical reality" that young fetuses often die prior to birth.[151] "Accordingly, the mere fact that a fetus aborts at some time after the woman carrying it is intentionally or unintentionally struck by another" does not conclusively prove that the battery caused fetal death.[152] Where the fetus was nonviable or immature, Justice Mosk stated that a prosecutor could hypothetically mislead the jury into convicting a defendant of feticide without establishing causation with certainty.[153]

The force of this argument, however, could be compromised by advances in medical technology that may be able to show conclusively the genesis of a miscarriage.[154] Furthermore, the vagueness arguments presented above, if successful, would only invalidate certain narrow aspects of the UVV; a comparable statute, specifying intent, viability, or causation standards, could be tailored to avoid these constitutional problems. Although opponents of the Act may raise due process challenges, procedural arguments such as these are unlikely to invalidate the central proposition of the Act: that fetuses can be treated as persons outside of the abortion context.

The Eighth Amendment's prohibition of cruel and unusual punishment may come closer to overruling the substance of the UVV.[155] The Supreme Court recognizes that punishment disproportionate to the crime committed constitutes cruel and unusual punishment.[156] The Court, in *Solem v. Helm*, listed three factors to determine whether a punishment is proportional to a crime.[157] A reviewing court should consider: the "gravity of the offense and the harshness of the penalty"; the "sentences imposed on other criminal defendants in the same jurisdiction"; and the "sentences for commission of the same crime in other jurisdictions."[158] While the Court held that these three elements should be considered in assessing proportionality, it emphasized that reviewing courts should

---

percent of all fertilizations do not end in live births." Barlow, *supra* note 17, at 493–94.

[151] *Davis*, 872 P.2d at 620 (Mosk, J., dissenting).

[152] *Id.*

[153] *See id.*

[154] *See* H.R. Rep. No. 107-42, pt. 1, at 2–3 (stating that "[t]he use of ultrasound, fetal heart monitoring, in vitro fertilization, and fetoscopy has greatly enhanced our understanding of the development of unborn children.").

[155] The Eighth Amendment reads, "[e]xcessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted." U.S. CONST. amend. VIII.

[156] *See* Solem v. Helm, 463 U.S. 277, 284 (1983) ("[T]he clause prohibits not only barbaric punishments, but also sentences that are disproportionate to the crime committed."). Some members of the Court refuse to recognize a proportionality requirement for punishment. *See, e.g.*, Harmelin v. Michigan, 501 U.S. 957, 965 (1991). In addition, some lower courts attempt to confine *Solem* to narrow circumstances. *See, e.g.*, U.S. v. Rhodes, 779 F.2d 1019, 1028 (4th Cir. 1985) (finding *Solem* to be applicable only in cases involving a life sentence without the possibility of parole). A proportionality principle nevertheless remains an active part of Eighth Amendment jurisprudence. *See* Barlow, *supra* note 17, at 490–91; *see also* Cooper Industries, Inc. v. Leatherman Tool Group, Inc., 121 S. Ct. 1678, 1684–85 (2001) (citing *Solem* favorably).

[157] *See Solem*, 463 U.S. at 291–92.

[158] *Id.*

grant broad deference to both legislatures and trial courts in their punishment and sentencing decisions.[159] Balancing these considerations in the context of the UVV indicates that the Act could survive an Eighth Amendment challenge; while allocating punishment as if a defendant had killed a born human being may be unduly harsh, the UVV's sentences are similar to those that have been applied and upheld in state feticide laws.

First, the criminal punishment provided under the Act may not be proportional to the gravity of the offense.[160] Regardless of moral or ethical beliefs, killing or causing bodily injury to an unborn child seems less severe, both physically and legally, than the same violence committed against a born human: a product of conception is significantly smaller, less developed, less stable, and less capable of living than a human being.[161] Furthermore, federal law holds that a fetus is less than a whole "person."[162] Abortion law specifically subordinates the potential life of a fetus to the privacy interest of its mother; a woman may choose to abort her pre-viable fetus without suffering penal consequences.[163] Even though states may proscribe consensual abortion after viability, under *Roe* and *Casey*, they must always make exceptions for the life and health of the mother.[164] The woman's life is, therefore, always treated as more important than her fetus's.[165] Similarly, most state fetal homicide statutes adopt different stages of gestation as their benchmarks for determining when fetal killing becomes homicide.[166] In so doing, these states implicitly recognize that terminating a pregnancy is not at all times equal to killing a human being.

The UVV, however, penalizes both intentional and unintentional crimes against the fetus as if a human being were the victim, notwithstanding state and federal precedent that holds that a fetus is not a full person.[167] Because a fetus is less than a person,[168] terminating a pregnancy must be a less grave offense than killing a born human. Thereby, the Act's penalty provisions are unduly harsh because they punish defendants as if they had committed the graver offense against a born person.

---

[159] *Id.*

[160] *Cf.* Barlow, *supra* note 17, at 503 (analyzing the harshness of California's sentencing scheme for fetal homicide).

[161] *See Davis*, 872 P.2d at 614–15 (Mosk, J., dissenting); *see also* Barlow, *supra* note 17, at 472, 496.

[162] *See Roe*, 410 U.S. at 158, 162; *see also Casey*, 505 U.S. at 913–14 (Stevens, J., concurring in part and dissenting in part).

[163] *See, e.g.*, *Casey*, 505 U.S. at 846 (recognizing the "right of the woman to choose to have an abortion before viability . . . . Before viability, the State's interests are not strong enough to support a prohibition of abortion").

[164] *See Roe*, 410 U.S. at 163–64; *see also Casey*, 505 U.S. at 846.

[165] *See* Barlow, *supra* note 17, at 502. Note that, by requiring exceptions for the mother's health in addition to her life, abortion law subordinates the fetus's existence even to an interest less than a human being's life.

[166] *See supra* notes 27–31 and accompanying text.

[167] *See* H.R. 503, §§ 2–3.

[168] *See supra* notes 98–99 and accompanying text.

Applying the second prong of *Solem*—comparing penalties under the UVV with those sentences imposed on similar defendants under federal law—is inconclusive. Because the Act represents the first time that federal law has penalized feticide, no similar prior sentencing scheme can be compared with the UVV's structure.[169]

The third factor in *Solem*, comparison to sentences imposed on similar defendants in other jurisdictions, reveals a wide range of sentencing patterns. On the harshest end of the scale,[170] California imposes the death penalty for fetal murder through its special circumstances statute.[171] When a defendant terminates a pregnancy during the commission of certain designated felonies, the state can also invoke the death penalty.[172] The UVV's sentencing structure is less harsh in comparison because it specifically prohibits the death penalty.[173]

The UVV is comparable to many state statutes in that it makes life imprisonment the harshest punishment available.[174] The UVV mandates life imprisonment in cases where a defendant intentionally kills a fetus with "malice aforethought."[175] Short of first degree feticide, the UVV imposes lower gradations of penalties in proportion to the level of intent manifested by the defendant.[176] In this regard, the UVV is similar to some criminal statutes, like Minnesota's, which criminalize feticide at all stages of gestation;[177] the Act, however may be arguably harsher than

---

[169] *Cf.* Barlow, *supra* note 17, at 504.

[170] *See Davis*, 872 P.2d at 620 (Mosk, J., dissenting).

[171] *See* CAL. PENAL CODE § 190.2 (West 1999). Because fetal murder is charged under the general murder statute, California courts have found that the state can charge a double murder special circumstance against a defendant who kills a fetus and at least one additional person or fetus. *See, e.g.*, People v. Bunyard, 756 P.2d 795 (Cal. 1988); *see also* CAL. PENAL CODE § 190.2(a)(3) (West 1999). This charge carries permission to request a death sentence. CAL. PENAL CODE § 190.2(a) (1999).

[172] *See* CAL. PENAL CODE §§ 189, 190.2(17) (West 1999); *see also Davis*, 872 P.2d at 619 (Mosk, J., dissenting).

[173] H.R. 503, §§ 2–3.

[174] *See*, H.R. 503, §§ 2–3. *See, e.g.*, CAL. PENAL CODE § 190.2(a) (West 1999); GA. CODE ANN. § 16-5-80 (1992); MINN. STAT. § 609.2661 (1998).

[175] *See* H.R. 503, §§ 2–3; *see also* 18 U.S.C. § 1111 (1994); *supra* text accompanying note 43–44.

[176] *See* H.R. 503, §§ 2–3. If the defendant lacked malice, for example, he would be charged only with manslaughter, which carries a maximum prison sentence of ten years. *See* 18 U.S.C. § 1112 (1994). Although the UVV does not have a *mens rea* or malice requirement, the provision covering intentional termination of a pregnancy incorporates the malice elements by reference to the applicable sections of the Federal Criminal Code, which themselves provide for gradations of punishment based on whether or not malice is present. *See* H.R. 503, §§ 2–3; *see also* 18 U.S.C. §§ 1111–1112 (1994).

[177] *See* MINN. STAT. § 609.2661–5 (1998). Minnesota similarly provides for different levels of punishment according to intent. First-degree murder carries a life sentence, *id.* § 609.2661, while second-degree murder merits no more than forty years imprisonment, *id.* § 609.2662, and third-degree murder no more than twenty-five years. *Id.* § 609.2663. First-degree manslaughter carries a penalty of no more than fifteen years imprisonment, a fine of no more than $30,000, or both, *id.* § 609.2664, while second-degree manslaughter lowers the longest prison term to ten years and the fine to $20,000. *Id.* § 609.2665.

feticide statutes like Georgia's, which do not impose any penalties at all until a later stage of pregnancy.[178]

The three factors laid out in *Solem* thus indicate that, while the UVV may impose penalties much harsher than the offense is grave, the Act would probably survive an Eighth Amendment attack. Its penalties are either proportional to or less harsh than those of state criminal statutes, which have been upheld in state courts of law.[179] In addition, because the *Solem* Court emphasized that deference to the legislature's determination of punishment is appropriate, Congress's passage of the UVV would accord some presumption of legitimacy to the Act.[180]

Although the UVV may well survive Fifth Amendment and Eighth Amendment challenges, it remains in conflict with abortion law. While the Act disclaims its power to affect abortion rights, the substance of the UVV appears to contradict the fundamental premises of abortion law—that the Fifth and Fourteenth Amendments do not include fetuses in the definition of "person"—by punishing violence against fetuses by third parties as harshly as violence against human beings. To avoid such a conflict and its potential threat to undermine the right to choose abortion, lawmakers should attempt to create alternatives to the UVV that do not contain the contentious elements present in the UVV.[181] Otherwise, while the Act may not legally affect the right to abortion, its rhetoric will likely color the abortion debate and the legal battles of the next century.

—Tara Kole
Laura Kadetsky

---

[178] *See* GA. CODE ANN. § 16-5-80 (1992).

[179] *E.g.*, *Bunyard*, 756 P.2d at 828–29. The defendant in *Bunyard* claimed that sentencing him to death for the murder of one person and one fetus under the state's double murder special circumstance provision violated the Eighth Amendment. *See id.* The court refused to entertain his claim, stating that, because California considered the killing of a fetus with malice aforethought to be murder under the general murder statute, and because murder was the gravest of crimes, feticide could be punished as the statute specified without offending the Constitution. *See id.* at 829–30.

[180] *See Solem*, 463 U.S. at 291–92.

[181] A better alternative to the UVV may be the Motherhood Protection Act ("MPA"), which was proposed and rejected as an amendment to the UVV in the 107th Congress. *See* 147 CONG. REC. E594 (daily ed. Apr. 24, 2001) (statement of Rep. Lofgren). This proposal provides penalties similar to the UVV for criminal defendants without compromising abortion jurisprudence by focusing on protecting the pregnant woman; the MPA makes it a separate federal crime to harm a pregnant woman. *See id.* Critics of the UVV support this substitute bill because it "avoids the issues of 'fetal rights' and 'fetal personhood . . . .'" *Id.* This bill could better deter criminal conduct against women without conflicting so markedly with established constitutional law.

THE BANKRUPTCY ABUSE PREVENTION AND CONSUMER PROTECTION ACT

The number of personal bankruptcies has nearly doubled in the past ten years: it rose from roughly 700,000 in 1990 to more than 1.2 million in 2000, passing through a peak of nearly 1.4 million in 1998.[1] Despite the prosperity of the 1990s, roughly one percent of American households declared bankruptcy each year.[2] In response to these figures, and to aggressive lobbying by banks and consumer lenders,[3] broad bipartisan majorities in both the House and Senate passed the Bankruptcy Abuse Prevention and Consumer Protection Act ("Bankruptcy Reform Act") early in the 107th Congress.[4] President Bush has indicated he will sign the bill,[5] which is virtually identical to legislation that President Clinton killed by pocket-veto in the 106th Congress.[6]

Proponents describe the Bankruptcy Reform Act as an effort to restore an ethic of personal responsibility among borrowers by forcing them to repay what debts they can.[7] At a time of impending recession, they argue, the bill will deter reckless borrowing, generating savings for lenders that will be passed on to consumers in the form of lower interest rates.[8] As prominent academics, congressional Democrats, and consumer

---

[1] *See* American Bankruptcy Institute, *U.S. Bankruptcy Filings 1980–2000 (Business, Non-Business, Total)*, *available at* http://www.abiworld.org/stats/1980annual.html (last visited Oct. 26, 2001).

[2] *See* Michelle J. White, *Why Don't More Households File for Bankruptcy?*, 14 J.L. ECON. & ORG. 205 (1998).

[3] *See* 145 CONG. REC. S14,067 (daily ed. Nov. 5, 1999) (statement of Sen. Feingold (D-Wis.) (indicating that "[a] very wealthy and powerful industry has pushed and pushed and pushed for this bill," and that "the members of the National Consumer Bankruptcy Coalition, an industry lobbying group made up of the major credit card companies such as Visa and MasterCard and associations representing the Nation's big banks and retailers, gave nearly $4.5 million in contributions to parties and candidates" during the 1998 election cycle); *see also* Kathleen Day, *House Passes Bankruptcy Limits: Measure Would Make it Harder for Consumers to Wipe Out All Debts*, WASH. POST, Mar. 2, 2001, at A1 ("Contributions to federal candidates and the political parties from finance and credit card companies during the 2000 campaign totaled $9.2 million .... Commercial banks' political contributions [totaled] $28.5 million.").

[4] H.R. 333, 107th Cong. (2001); S. 420, 107th Cong. (2001). The House Bill passed on March 1, 2001 by a margin of 306 to 108, receiving 93 Democratic votes in favor and no Republican "nay" votes. *See* CONG. REC. H600–01 (daily ed. Mar. 2, 2001). The Senate then passed the bill with minor amendments on July 17, 2001 by a margin of eighty-two to sixteen. *See* 147 CONG. REC. S7742 (daily ed. July 18, 2001).

[5] *See* Day, *supra* note 3, at A1.

[6] *See* HOUSE COMM. ON THE JUDICIARY, BANKRUPTCY ABUSE PREVENTION AND CONSUMER PROTECTION ACT OF 2001, H.R. REP. NO. 107-3, at 3 (2001) [hereinafter HOUSE JUDICIARY COMMITTEE REPORT].

[7] *See, e.g.*, *id.* at 2 ("The purpose of the bill is to improve bankruptcy law and practice by restoring personal responsibility and integrity in the bankruptcy system ...."); 147 CONG. REC. H518 (daily ed. Mar. 1, 2001) (statement of Rep. Armey (R-Tex.), House Majority Leader) ("This bill is about the character of a Nation and [whether] the Nation's laws have a character of the Nation's people.").

[8] *See, e.g.*, 147 CONG. REC. S1807 (daily ed. Mar. 5, 2001) (statement of Sen. Grassley (R-Iowa)) ("With the possibility of the economy slowing right now, we need to at this time

advocates have noted, however, the bill carries a strong pro-creditor bias that threatens to undermine its purported objectives.[9] Since the bill does little to curb the lending practices that have made possible Americans' extraordinary debts, there is reason to believe it will succeed neither in reducing bankruptcies nor lowering interest rates.[10]

The main thrust of the reform bill is to provide less generous options to debtors who file for bankruptcy. Under current law, debtors may choose between two personal bankruptcy options: asset liquidation under Chapter Seven[11] of the bankruptcy portion of the United States Code, and debt readjustment under Chapter Thirteen.[12] Any debtor may seek the protection of either chapter, regardless of whether the debtor is insolvent.[13] Chapter Seven permits the filer to discharge all debts, with certain exceptions such as mortgages, student loans, fraudulently incurred debts, and some child support and alimony obligations.[14] In return for the discharge, however, the debtor must accept the liquidation of all assets, excepting certain exempt property such as the debtor's homestead.[15] By

---

fix a bankruptcy system that inflates interest rates and threatens to make the slowdown even worse.").

[9] *See* Kathleen Day, *Foes of Bankruptcy Bill Point Finger at Credit Card Issuers*, WASH. POST, Feb. 28, 2001, at E1; *see also* HOUSE COMM. ON THE JUDICIARY, BANKRUPTCY ABUSE PREVENTION AND CONSUMER PROTECTION ACT OF 2001: DISSENTING VIEWS, H.R. REP. NO. 107-3, at 455 (2001) [hereinafter JUDICIARY COMMITTEE DISSENTING VIEWS] ("We oppose the bill because it is likely to harm low income consumers, women and children reliant on alimony and child support, and employees of troubled businesses, among other vulnerable groups."); 147 CONG. REC. H518 (daily ed. Mar. 1, 2001) (statement of Rep. Conyers (D-Mich.), Ranking Member, House Judiciary Comm.) (calling House Bill 333 "a bill that massively tilts the playing field in favor of creditors and against the interests of ordinary consumers and workers.").

[10] *See, e.g.*, JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 478 (criticizing the Bankruptcy Reform Act for "ignor[ing] the transgressions of the credit industry," and noting that "the overwhelming weight of authority establishes that it is the massive increase in consumer debt ... which has brought about the increases in consumer filings."); *infra* text accompanying notes 57–88.

[11] 11 U.S.C. §§ 701–766 (1994 & Supp. V 1999).

[12] *Id.* §§ 1301–1330.

[13] *See id.* § 707(b) (stating that a court may deny access to Chapter Seven only if it finds that "the granting of relief would be a substantial abuse of the provisions of this chapter."). Commentators have disputed the extent to which this provision protects creditors from bankruptcies by debtors who could afford to pay their debts. *Compare* SENATE JUDICIARY COMMITTEE, BANKRUPTCY REFORM ACT OF 1999, S. REP. NO. 106-49, at 6 (1999) (indicating that, since the statute provided no definition of "substantial abuse," courts have generally held excess income is only one factor to be considered) *with* Carl Felsenfeld, *Denial of Discharge for Substantial Abuse: Refining—Not Changing—Bankruptcy Law*, 67 FORDHAM L. REV. 1369 (1999) (noting a split in Circuits between two tests for "substantial abuse," the "excess income" test and the "totality of the circumstances" test, but arguing that in practice "courts routinely apply only an excess income test"). A finding of "substantial abuse" is possible only upon motion of the court or of the United States Trustee, a Department of Justice official who manages the liquidation of the debtor's assets; creditors and interested parties may not request such a finding. *See* 11 U.S.C. at § 707(b) (Supp. V 1999).

[14] *See* 11 U.S.C. §§ 523, 727 (1994 & Supp. V 1999).

[15] *See id.* § 522. The debtor may choose between exemptions provided by federal or state law. *See id.* § 522(b). This exemption system is controversial. *See infra* note 46.

contrast, debtors who file under Chapter Thirteen may retain most assets, but must develop a plan to commit their disposable income to repayment of debts over a three to five year period,[16] at the end of which the court discharges any remaining debt.[17]

The Bankruptcy Reform Act would abolish the debtor's voluntary choice of chapter, replacing it with a means test designed "to ensure that debtors repay creditors the maximum they can afford."[18] The bill would require all bankruptcy filers to complete the test, though in general only those with a family income above the state median would be denied access to Chapter Seven.[19] Under the bill's test, the debtor's monthly income (defined as the monthly average of all income received during the previous six months)[20] multiplied by 60 (the number of months in five years, the length of a typical Chapter Thirteen repayment plan under the bill)[21] is compared to the sum of the following five figures: (1) total "priority" debts, such as family support obligations; (2) scheduled payments on secured debts over the next five years; (3) arrears on secured debts; (4) monthly living expenses calculated on the basis of Internal Revenue Service guidelines, multiplied by 60; and (5) an allowance for the administrative costs of a Chapter Thirteen plan.[22] If the debtor's income exceeds this sum by either (1) an amount equal to twenty-five percent of remaining unsecured debts (unless the debtor could pay no more than $100 a month toward these debts) or (2) $10,000, whichever is lower, then the debtor faces a presumption of abuse which he or she may rebut only by establishing that, due to "special circumstances," there is "no reasonable alternative" to an adjustment in the debtor's income or expenditure.[23]

While the means test is the Bankruptcy Reform Act's centerpiece,[24] it is not the only new restriction the bill would impose on bankruptcy filing. The bill would also, for one, require debtors to obtain "credit

---

[16] *See* 11 U.S.C. § 1322 (1994).

[17] *See id.* § 1328 (1994 & Supp. V 1999).

[18] HOUSE JUDICIARY COMMITTEE REPORT, *supra* note 6, at 2.

[19] *See* H.R. 333, § 102(a).

[20] *See* H.R. 333, § 102(b). The calculation of monthly income excludes Social Security benefits and reparation payments for victims of war crimes and crimes against humanity; however, it includes all other sources of income whether or not they are taxable. *See id.*

[21] *See* 511 U.S.C. § 1322 (1994). Whereas three years is the length of the typical Chapter Thirteen plan under current law, the Bankruptcy Reform Act would make five years the norm in an increased number of cases. *See* H.R. 333, § 318.

[22] *See* H.R. 333, § 102(a).

[23] *See id.* Chapter Seven access would be presumptively denied to a debtor with above-average income who could pay all scheduled payments and arrears on priority obligations plus at least $6,001 (more than $100 per month over five years) towards a $24,000 credit card debt (four times the minimum payable amount). Chapter Seven would also be unavailable to a debtor who could pay all priority obligations plus $10,000 of a $10,001 credit card debt.

[24] *See* HOUSE JUDICIARY COMMITTEE REPORT, *supra* note 6, at 2 (calling the means test the bill's "heart").

counseling" before filing.[25] Under this provision, debtors could receive bankruptcy relief only upon completion of "an instructional course concerning personal financial management" given by an approved non-profit counseling agency.[26] For another, the bill would toughen the presumptions against discharge of debt obtained shortly before filing. Current law permits creditors to challenge the discharge of any debt of $1,150 or more for "luxury goods and services" that the debtor obtained from a single creditor within sixty days of filing.[27] The Bankruptcy Reform Act would reduce the amount for such presumptively non-dischargeable debts to $250, while increasing the relevant time-span from sixty to ninety days before bankruptcy.[28] Similarly, whereas current law prevents the discharge of debts from cash advances amounting to more than $1,150 if they are incurred within sixty days of filing,[29] the Bankruptcy Reform Act would lower that minimum amount to $750, while also shortening the relevant time-period to seventy days.[30] In addition, the bill would restrict debtors' rights to stay debt payment by declaring bankruptcy.[31] There is evidence that some debtors make strategic use of the automatic stay afforded to filers under current law;[32] some debtors apparently file successive Chapter Thirteen motions with no intention of completing a plan and obtaining a discharge.[33] The Bankruptcy Reform Act would curtail such "abusive serial filing" by requiring a showing of good faith before granting a stay of more than thirty days to debtors who had filed and had their case dismissed within the previous year.[34]

Because such changes to the mechanics of bankruptcy under the Bankruptcy Reform Act could alter the relative advantages for competing creditor groups, the bill would adjust the treatment of certain types of debts in bankruptcy. Under current law, Chapter Seven filings are generally more advantageous for creditors holding secured debt. Chapter

---

[25] H.R. 333, § 106. For critical analysis of the view that a credit counseling requirement would benefit debtors, see Howard B. Hoffman, *Consumer Bankruptcy Filers and Pre-Petition Consumer Credit Counseling: Is Congress Trying to Place the Fox in Charge of the Henhouse?*, 54 BUS. LAW. 1629, 1632 (1999) (indicating that, because existing credit counseling agencies are funded by "voluntary contributions" from creditors, such agencies may encourage debtors to undertake burdensome repayment budgets).

[26] H.R. 333, § 106(b)(3) (adding the counseling requirement to Chapter Seven); *id.* § 106(c) (adding the requirement to Chapter Thirteen).

[27] 11 U.S.C. § 523(a)(2)(C) (LEXIS through 2001 legislation).

[28] *See* H.R. 333, § 310.

[29] *See* 11 U.S.C. § 523(a)(2)(C) (LEXIS through 2001 legislation).

[30] *See* H.R. 333, § 310.

[31] *Id.* § 302.

[32] *See* 11 U.S.C. § 362 (1994 & Supp. V 1999).

[33] *See* Elizabeth Warren, *A Principled Approach to Consumer Bankruptcy*, 71 AM. BANKR. L.J. 483, 502–03 (1997).

[34] H.R. 333, § 302. Debtors who file more than twice within a year would have to show good faith by clear and convincing evidence. *See id. See generally* Scott F. Norberg, *Consumer Bankruptcy's New Clothes: An Empirical Study of Discharge and Debt Collection in Chapter 13*, 7 AM. BANKR. INST. L. REV. 415, 427 (1999) (describing the abusive serial filing provisions).

Seven debtors often agree to reaffirm secured debts so as to keep the underlying collateral, such as an automobile or appliance being purchased in installments. Many even accept agreements that require them to pay an outstanding balance that exceeds the value of the collateral property.[35] Meanwhile, unsecured lenders, such as credit card issuers, tend to lose out when assets are liquidated under Chapter Seven because unsecured debt is generally a low priority for repayment.[36]

By contrast, Chapter Thirteen is more advantageous to unsecured lenders. As an incentive for debtors to file under Chapter Thirteen rather than Seven, secured debts on assets other than the debtor's home may be "stripped down" from the amount of the outstanding balance to the replacement value of the property.[37] Thus, whereas secured creditors may often obtain reaffirmation of the full value of their lien against a Chapter Seven debtor, the value of secured debts may be substantially reduced if the debtor files under Chapter Thirteen. Meanwhile, both secured and unsecured debt are included in the repayment plan, so unsecured lenders may receive repayment of debts that would have been discharged under Chapter Seven.[38]

---

[35] *See* Warren, *supra* note 33, at 499 (noting that "more than forty percent of the debtors have been willing to sign on [to reaffirmation agreements]" even though those agreements typically require them to repay not only the secured portion of the debt, but also the unsecured portion, that is, the portion in excess of the value of the collateral). Reaffirmation agreements are permitted by 11 U.S.C. § 524(c) (1994 & Supp. V 1999). In an empirical study, Marianne Culhane and Michaela White found that twenty-eight percent of Chapter Seven debtors had "one or more reaffirmation agreements in the court file." Marianne B. Culhane & Michaela M. White, *Debt after Discharge: An Empirical Study of Reaffirmation*, 73 AM. BANKR. L.J. 709, 720 (1999). These figures might understate the level of reaffirmation, since there is evidence that many creditors obtained illegal "rogue reaffirmations" requiring debtors to repay loans without reporting the agreement to the court. *Id.* at 717–18. Many debtors in their sample also appeared to have retained assets through "ride-through" arrangements that permitted them to continue payments on items purchased on credit with the creditor retaining a right to repossess the item. *Id.* at 713. Culhane and White conclude that many Chapter Seven filers reaffirmed more debt than they could afford to pay; moreover, many reaffirmed debts on household items that were "unsecured for all practical purposes" since "the collateral was unlikely to be repossessed." *Id.* at 764.

[36] *See* 11 U.S.C. §§ 507, 726 (1994 & Supp. V 1999). Unsecured lenders have sometimes succeeded in obtaining reaffirmations of debt in return for promises of credit following bankruptcy. *See* Culhane and White, *supra* note 35, at 730; *see also* Scott F. Norberg, *supra* note 34, at 421 ("While unsecured creditors rarely realize any payment through liquidation of unencumbered, non-exempt property, they routinely receive at least some repayment through reaffirmations and non-dischargeability determinations.").

[37] *See* 11 U.S.C. § 1322(b)(2) (1994) (permitting the court to "modify the rights of secured claims, other than a claim secured only by a security interest in real property that is the debtor's principle residence"). Courts have interpreted this provision in conjunction with § 506(a), which provides that debt in excess of the value of the collateral may be treated as unsecured debt if another provision of the code permits modification of the lien. *See, e.g.*, *In re* Byington, 197 B.R. 130, 133 (Bankr. D. Kan. 1996); *see also* Corinne Ball & Jacqueline B. Stuart, *The Battle Over Bankruptcy Law for the New Millennium*, 55 BUS. LAW. 1487, 1496–97 (2000).

[38] *See generally* Norberg, *supra* note 34, at 424–26.

Forcing debtors to file under Chapter Thirteen could disadvantage secured creditors if it eliminated reaffirmation opportunities that would have arisen under Chapter Seven. The Bankruptcy Reform Act compensates for this effect by restricting the debtor's opportunity to strip down undersecured debts in Chapter Thirteen. For automobile purchase loans made within five years of bankruptcy or for other personal property purchased within one year, debtors would no longer be permitted to bifurcate the replacement value from the remaining unsecured portion of the loan; the bill would instead require debtors to include the full value of such creditor liens in their repayment plan.[39] Further, the bill would require reaffirmation agreements to include extensive information about the debtor's rights,[40] and would instruct United States Attorneys and the Federal Bureau of Investigation to take responsibility for enforcing laws against abusive reaffirmation practices.[41]

Finally, the Bankruptcy Reform Act would amend the Truth-in-Lending Act ("TILA")[42] to require new disclosures from creditors. In particular, it would require credit card billing statements to disclose information on the amount of time required to pay off a hypothetical balance by making minimum monthly payments.[43] Card issuers would be required to provide a toll-free telephone number from which card-holders could obtain information on the amount of time required to pay off their own account balance with minimum payments.[44] Additionally, credit card solicitations and applications would be required to include information "clearly and conspicuously" on the mechanics and duration of introductory "teaser" interest rates.[45]

In sum, the Bankruptcy Reform Act would restrict access to Chapter Seven, limit the debt discharge, impose new burdens on bankruptcy filers, and eliminate strip-down, while offering consumers the benefit of additional disclosures from consumer creditors.[46] With the exception of the

---

[39] *See* H.R. 333, § 306. House Bill 333, section 304 also bars debtors and creditors from agreeing to let secured claims "ride-through" bankruptcy, as they may under current law. *See* Culhane & White, *supra* note 35, at 718–20 (describing "ride-through" options under current law).

[40] *See* H.R. 333, § 203(a).

[41] *See id.* § 203(b); *See also* Culhane & White, *supra* note 35, at 717–18 (describing abusive reaffirmation practices by creditors).

[42] 15 U.S.C. § 1637 (1994).

[43] *See* H.R. 333, § 1301(a). Credit card lending arrangements often involve "negative amortization"; that is to say, credit card lenders often set minimum monthly payments below interest charges, with the result that a debtor's total outstanding debt may increase while the debtor makes the minimum payments. TERESA A. SULLIVAN, ELIZABETH WARREN, & JAY LAWRENCE WESTBROOK, THE FRAGILE MIDDLE CLASS: AMERICANS IN DEBT 247–48 (2000).

[44] *See* H.R. 333, § 1301(a).

[45] *Id.* § 1303.

[46] The Bankruptcy Reform Act is a complex and detailed piece of legislation. This Essay discusses only the most significant and controversial changes the bill would make to current bankruptcy law. *See generally* Ball & Stuart, *supra* note 37. One controversial fea-

TILA amendments, all of these changes "move in the same direction": they strengthen the position of creditors while weakening that of debtors.[47]

The rationale for this package of reforms is twofold. First, proponents of the bill attribute the explosive increase in bankruptcy filing to abuse of current law by unscrupulous debtors.[48] Easy access to Chapter Seven, the argument goes, has eroded the "stigma" traditionally associated with bankruptcy, leading to the use of the bankruptcy system as a tool of financial planning, rather than as a last resort.[49] Bankruptcy laws, it follows, must be toughened to curb this practice.[50] Second, proponents argue that reform will have beneficial effects for responsible borrowers

---

ture of current law that the bill may *not* change significantly bears mention. Under current law, debtors may exempt assets from the bankruptcy estate according to either federal or state law. *See* 11 U.S.C. § 522 (1994 & Supp. V 1999). Some wealthy debtors have used this feature of bankruptcy law strategically, changing their state of residence before bankruptcy so as to gain a more generous exemption. *See* G. Marcus Cole, *The Federalist Cost of Bankruptcy Exemption Reform*, 74 AM. BANKR. L.J. 227, 229 (2000) (referring to "'high profile' cases in which wealthy individuals move to states with generous exemption regimes in order to obtain a discharge while shielding vast portions of their assets from creditors."). For instance, a debtor who moves from New Jersey, which provides no state-law homestead exemption, to Florida, which provides an unlimited homestead exemption, may shield substantial assets from creditors by purchasing an expensive home. *See* Ball & Stuart, *supra* note 37, at 1499–1500. Critics argue that this opportunity for pre-bankruptcy planning by the wealthy undermines public confidence in the fairness and legitimacy of the system. *See, e.g.*, Elizabeth Warren, *supra* note 33, at 493–94. While the Senate version of the Bankruptcy Reform Act would impose a cap of $125,000 on homestead exemptions, *see* Senate Bill 420 § 308, the House bill would make no such change, and President Bush has indicated that he opposes the Senate language. *See Senate Votes to Invoke Cloture on Motion to Proceed to Bankruptcy*, CONG. DAILY, July 12, 2001, *available at* 2001 WL 24848595.

[47] *Bankruptcy Reform: Joint Hearing Before the Subcomm. on Commercial and Administrative Law of the House Comm. on the Judiciary and the Subcomm. on Administrative Oversight and the Courts of the Senate Comm. on the Judiciary*, 106th Cong. 103 (1999) [hereinafter *Bankruptcy Reform Hearing*] (statement of Elizabeth Warren, Professor, Harvard Law School).

[48] *See, e.g., id.* (statement of Todd Zywicki, Professor, George Mason University School of Law) ("Bankruptcy is a moral as well as an economic act. There is a conscious decision not to keep one's promises. It is a decision not to reciprocate a benefit received, a good deed done on the promise that you will reciprocate. Promise-keeping and reciprocity are the foundation of an economy and healthy civil society."), *quoted in* HOUSE JUDICARY COMMITTEE REPORT, *supra* note 6, at 6.

[49] S. REP. NO. 106-49, at 2–3 ("A decreased moral stigma associated with bankruptcy means that filing for bankruptcy is not viewed as a last resort for financially troubled Americans who need debt forgiveness . . . . Individuals who would have struggled to meet their financial obligations in the past are filing bankruptcy today in record numbers.") (footnotes omitted); *see also* 146 CONG. REC. S50 (daily ed. Jan. 26, 2000) (statement of Sen. Hatch (R-Utah)) ("Not long ago in our Nation's past, there was an expectation that people should repay what they have borrowed. Hand in hand with this expectation was a stigma that attached to those who filed bankruptcy . . . . Our current system, I am sorry to say, allows some people who are able to repay their debts to avoid doing so."). *See generally supra* note 7.

[50] *See, e.g.*, Press Release, Sen. Chuck Grassley, Memorandum Re: S. 420, The Bankruptcy Reform Bill (Mar. 14 2001) ("There's plenty of blame to go around for the 100 percent increase in the number of bankruptcy filings in the last ten years. Part of the problem is bankruptcy laws that are too lax."), *at* http://www.senate.gov/~grassley/releases/2001/p01r3-14d.htm.

who pay back their loans. Under current law, responsible borrowers pay high interest rates to subsidize losses from the discharged debt of their less responsible peers.[51] A reform that limits access to debt discharge should lower interest rates, reformers argue, since it would prevent reckless borrowers from shifting the costs of their activity onto other borrowers.[52] Lower interest rates, in turn, would benefit the economy by facilitating responsible spending and investment.[53]

Opponents of the Bankruptcy Reform Act contest both rationales given in support of the reform bill. They argue, first, that lenders, not borrowers, are to blame for bankruptcy increases, since their lending practices have fostered the indebtedness that places households at risk of bankruptcy.[54] Opponents also question whether bankruptcy reform will lower interest rates for responsible borrowers.[55] They suggest, rather, that commercial lenders will be the principal beneficiaries of the bill, since it will afford them increased opportunities for debt collection.[56]

The first ground for disagreement relates to the causes of bankruptcy. On that count, one fact is clear: Americans have taken on unmanageable levels of debt in recent years. Indeed, the data on this point are overwhelming: total household debt in the United States increased from sixty-five percent of total income to eighty-one percent between 1980 and 1994.[57] During roughly the same period, households increased their total home mortgage and consumer installment debts by more than 400

---

[51] *See, e.g.*, HOUSE JUDICIARY COMMITTEE REPORT, *supra* note 6, at 5 (indicating that "financial losses attributable to bankruptcy filings in 1997 exceeded $44 billion"); 147 CONG. REC. H134 (daily ed. Jan. 31, 2001) (statement of Rep. Gekas (R-Pa.)) (indicating the $44 billion in losses from bankruptcy "equal[s] more than $400 per household"); 147 CONG. REC. S1811 (daily ed. Mar. 5, 2001) (statement of Sen. Sessions (R-Ala.)) ("When somebody fails to pay what they owe, . . . what happens? It drives up the cost of . . . people's business. They have to raise the charges on the honest people who pay them.").

[52] *See, e.g.*, 147 CONG. REC. S1807 (daily ed. Mar. 5, 2001) (Statement of Sen. Grassley (R-Iowa)) ("The result of the bankruptcy crisis is that hard-working, law-abiding Americans have to pay higher prices for goods and services. [The Bankruptcy Reform Act] makes it harder for individuals who can repay their debts to file for bankruptcy under chapter 7 where their debts are wiped away. This would lessen the upward pressure on interest rates and higher prices.").

[53] *See, e.g., id.* ("Bankruptcy reform will help our economy through lower interest rates.")

[54] *See, e.g.*, 147 CONG. REC. S2028 (daily ed. Mar. 8, 2001) (statement of Sen. Durbin (D-Ill.)) ("[Proponents of the Bankruptcy Reform Act] argued that the people who were filing for bankruptcy had forgotten the moral stigma of declaring bankruptcy in America . . . . Shouldn't the moral stigma be on the conscience of these lenders who have dragged these poor unsuspecting people into a situation where they have no hope and nowhere else to turn?").

[55] *See, e.g.*, JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 458 ("[W]e have never received any evidence that the credit card industry likely would pass on any of the 'savings' from bankruptcy law changes to individual consumers.").

[56] *See, e.g., id.* at 460 ("[The Bankruptcy Reform Act] would institute a number of major changes to consumer bankruptcy . . . that are designed to increase pay-outs to non-priority unsecured creditors, particularly credit card companies, as well as certain secured lenders, especially those extending credit for automobile loans.").

[57] SULLIVAN ET AL., *supra* note 43, at 18

percent.[58] Thus, "real consumer debt has risen dramatically over a long period during which real incomes for many people have stayed the same or declined."[59] In addition, according to one survey, bankruptcy filers in 1997 had an average debt-to-income ratio of 2.76, with the average burden of non-mortgage debt totaling 1.87 times annual income.[60] During 2000, the American savings rate was negative for the first time since the Great Depression.[61]

The explanation for these trends is not obvious, however. On the one hand, some research supports the reformers' claim that a shift in borrowers' attitudes is to blame. For instance, one recent study concludes that "the explosion in bankruptcy filings is in substantial part attributable to a shift in social norms," not legal and economic variables.[62] In the authors' view the most plausible explanation for increasing debt is "a decline in social sanctions"—that is, in stigma associated with bankruptcy.[63] Media reports of debtors filing for bankruptcy without compunction also suggest that social attitudes toward bankruptcy have changed.[64] Moreover, other research shows that increases in the rate of bankruptcy filing have correlated strongly with increased expenditures on legal advertising following a 1977 Supreme Court decision[65] that struck down restrictions on such advertising on free speech grounds.[66] While "it is difficult to establish a causal relationship,"[67] lawyer advertisements for bankruptcy services may have contributed to a decline in social stigma by presenting bankruptcy filing in a more positive light.[68]

On the other hand, declining stigma does not appear to be a complete explanation. Some research indicates that about fifteen times as many American households could benefit from bankruptcy as actually file.[69] Socio-cultural deterrents may afford at least a partial explanation

---

[58] *Id.*

[59] *Id.*

[60] *Id.* at 70–71 tbl.2.5.

[61] *See* Dolores Kong, *Many Seen at Risk in a Downturn: Personal Debt Soars While Savings Plunge*, BOSTON GLOBE, Mar. 20, 2001, at A1.

[62] F.H. Buckley & Margaret F. Brinig, *The Bankruptcy Puzzle*, 27 J. LEGAL STUD. 187, 188 (1998).

[63] *Id.* at 206.

[64] *See, e.g.*, Kim Clark, *Why So Many Americans Are Going Bankrupt?*, FORTUNE, Aug. 4, 1997, at 24; *see also* Judge Edith H. Jones & Todd J. Zywicki, *It's Time for Means-Testing*, 1999 BYU L. REV. 177, 215–21 (1999) (quoting in part Clark, *supra*, at 24–25). *But see* SULLIVAN ET AL., *supra* note 43, at 32 ("Arguments that the stigma attached to bankruptcy has declined are typically made by journalists who are unable to find any bankruptcy debtors willing to be interviewed for the record and by prosperous economists who see bankruptcy as a great bargain.") (footnote omitted).

[65] *Bates v. State Bar of Arizona*, 433 U.S. 350 (1977).

[66] *See* Diane Ellis, *The Influence of Legal Factors on Personal Bankruptcy Filings*, BANK TRENDS 98–103 (FDIC, Division of Insurance), Feb. 1998, *available at* http://www.fdic.gov/bank/analytical/bank.

[67] *Id.* at 9.

[68] *See id.* at 2.

[69] *See* White, *supra* note 2, at 205.

for this anomaly.[70] Indeed, some commentators doubt whether the stigma of bankruptcy has declined significantly.[71] Furthermore, changes in the culture and regulation of consumer lending appear to have played some role in bankruptcy increases. Bankruptcy levels have closely tracked overall levels of consumer debt,[72] which have risen rapidly with the burgeoning of the consumer credit industry beginning in 1978.[73] In that year, the Supreme Court made possible the modern credit card industry by permitting lenders to charge the highest interest rate available in any state to credit card borrowers throughout the country.[74] Credit was once all but unavailable for consumers who were too risky for loans within the regulated interest rate of their state.[75] Today, however, lenders bombard consumers across the United States with some three billion credit card solicitations annually—forty-one per household—spending roughly $100 in solicitation costs for each new card-holder.[76] Unlike traditional lenders, credit card companies do not base lending decisions on a careful analysis of the borrower's creditworthiness; rather, they extend credit on an "actuarial basis," targeting entire demographic groups and charging interest rates as high as twenty-four percent to cover the risk that some borrowers

---

[70] *Cf.* SULLIVAN ET AL., *supra* note 43, at 32 ("Discharging debts that were honestly incurred seems the antithesis of middle-class morality. Public identification as a bankrupt debtor is embarrassing at best, devastating at worst. Bankrupt debtors have told us of their efforts to conceal their bankruptcy."). *But see* White, *supra* note 2, at 229 (postulating that debtors do not file "because they obtain the benefits of bankruptcy without bearing the costs of filings, since creditors do not attempt to collect.").

[71] *See, e.g.*, SULLIVAN ET AL., *supra* note 43, at 32.

[72] *See* JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 478 (noting that "there is an almost perfect correlation between the increasing amount of consumer debt and the number of consumer bankruptcy filings"); *see also* SULLIVAN ET AL., *supra* note 43, at 129 ("Bankruptcy filing rates and consumer debt to income ratios rise and fall together over time."); Paul C. Bishop, *A Time Series Model of the U.S. Personal Bankruptcy Rate*, BANK TRENDS 98–101, (FDIC, Division of Insurance), Feb. 1998 (estimating the influence of consumer indebtedness and business cycle activity on the bankruptcy rate and concluding that "approximately two-thirds of the increase in bankruptcies can be explained by these two factors alone"), *available at* http://www.fdic.gov/bank/analytical/bank/bt_9801.pdf (last visited Jan. 1, 2002).

[73] *See* JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 8, at 478 (chart shows rapid increase in consumer debt levels from the 1970s to the present).

[74] *Marquette Nat. Bank of Minneapolis v. First Omaha Serv. Corp.*, 439 U.S. 299 (1978) (holding that usury laws regulating interstate lending are a policy matter for Congress to decide). *See generally* Dianne Ellis, *The Effect of Consumer Interest Rate Deregulation on Credit Card Volumes, Charge-Offs, and the Personal Bankruptcy Rate*, BANK TRENDS 98–105 (FDIC, Division of Insurance), Mar. 1998 (arguing that *Marquette* "fundamentally altered the market for credit card loans in a way that significantly expanded the availability of credit and increased the average risk profile of borrowers," leading to "a substantial expansion in credit card availability, a reduction in average credit quality, and a secular increase in personal bankruptcies."), *available at* http://www.fdic.gov/bank/analytical/bank/bt_9805.html (last visited Jan. 1, 2002).

[75] *See* Ellis, *supra* note 74, at 4 ("[I]n a regime of restrictive usury ceilings, where lenders' income potential was limited, lenders extended credit only to higher-quality borrowers, and poorer quality borrowers were shut out of the market.").

[76] *See* SULLIVAN ET AL., *supra* note 43, at 135.

in the group will be unable to pay.[77] Thus, credit is widely and easily available to consumers, regardless of creditworthiness. What is more, consumers may incur debts incrementally by charging purchases to their card without ever having to apply deliberately for a loan.[78] The result is that consumers have taken on more debt than ever before, with personal bankruptcy rates rising proportionally.[79] Aggressive efforts by lenders to extend credit card usage to new demographic groups, including low-income borrowers in the "subprime" market,[80] suggest that the rise in consumer indebtedness will be an ongoing trend.[81]

In addition to the increasing availability of credit, structural economic changes may be inducing households to take on more debt and increase their risk of bankruptcy. Increasing income inequality has produced a situation in which more families must borrow in order to maintain a "middle-class" lifestyle.[82] The ill effects of this trend are compounded by the fact that rising divorce rates and decreasing job security have increased the risk that families will undergo financial shocks that cause them to fall behind in debt payments.[83] Meanwhile, the combination of consumer advertising pressures with a deeply ingrained culture of "rising expectations"—the presumption that household welfare will improve as careers advance—makes it difficult for many Americans to save funds. Their lack of savings, in turn, leaves Americans ill-prepared for future calamities, let alone the belt-tightening that may be required in the case of more mundane problems such as diminished salary, layoff, temporary unemployment, or divorce.[84] Thus, the consumer credit industry has come to function as a form of private wage insurance which provides emergency funds to households in distress.[85] In this capacity, the industry finances the losses of families who go bankrupt by charging high interest

---

[77] *Id.* at 246–47. Eighteen percent is the average interest rate, but credit card issuers often raise the rate to twenty-four percent or higher for borrowers who fall behind in payments. *Id.* at 18–19.

[78] *See id.* at 130.

[79] A comparison between Canada and the United States suggests that credit-card lending has played a substantial role in consumer debt and bankruptcy increases. Modern credit cards were introduced in the mid-1960s in both the United States and Canada, but states regulated interest rates in the United States until the 1978 *Marquette* decision, whereas Canadian interest rates have been deregulated since at least 1886. *See* Ellis, *supra* note 74, at 9. The Canadian and American personal bankruptcy rates have both risen dramatically in the past twenty years, yet the Canadian rate rose immediately following the introduction of credit cards—rising by 340% between 1966 and 1976—while the American rate took off only after the *Marquette* decision made widespread marketing of credit cards profitable in the United States. *See id.* at 9–10; *see also* JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 479.

[80] Timothy L. O'Brien, *Lowering the Credit Fence: Big Players are Jumping Into the Risky Loan Business*, N.Y. TIMES, Dec. 13, 1997, at D1.

[81] *See* Ellis, *supra* note 74, at 10.

[82] *See* SULLIVAN ET AL., *supra* note 43, at 28–33.

[83] *See id.* at 75–107, 172–98.

[84] *See id.* at 22–26.

[85] *See id.* at 138.

rates to those who manage to recover and repay some or all of their debts.[86]

In the view of critics of the Bankruptcy Reform Act, the role of structural economic trends and creditor behavior in the recent bankruptcy boom argues against the bill's focus on discouraging bankruptcy by strengthening creditor protections. The dissenters on the House Judiciary Committee contend that "the vast weight of the data and studies contradict [sic] the proponents' rationales and instead shows that non bankruptcy law factors are the root cause of increased bankruptcy law filings."[87]

Reforms to benefit creditors are inappropriate, critics argue, for they benefit credit card lenders and other creditors who have helped create the bankruptcy crisis while burdening "vulnerable groups" that find themselves in debt as a result of financial and economic forces beyond their control.[88]

Most Senators and Representatives, however, appear to have been more persuaded by the contrary view that deterrents should be directed at borrowers, not lenders, notwithstanding any role lenders have played in fostering current bankruptcy levels. It may be that structural economic changes are the causes of bankruptcy. Nevertheless, the argument goes, if consumers are using the bankruptcy system to avoid the consequences of irresponsible purchases or to support unrealistic financial expectations, then it is appropriate for Congress to encourage saving and restraint by sending a "moral signal" that debts must be repaid.[89] For many reformers, the prospect of an economic recession adds force to this view since the risk of bankruptcy will likely increase for many households during hard times.[90] As for the concerns about vulnerable groups, the bill's means test purports to address them by ensuring that the bill adversely affects only those debtors who fail the means test and do not "need" Chapter Seven relief.[91] Thus, proponents have been able to present the bill as "balanced," notwithstanding critics' concerns about its generosity to creditors.[92]

---

[86] *See id.*

[87] JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 457.

[88] *Id.* at 455.

[89] Jones & Zywicki, *supra* note 64, at 191.

[90] *See, e.g.*, 147 CONG. REC. S1806 (daily ed. Mar. 5, 2001) (statement of Sen. Grassley (R-Iowa)) ("If we hit a recession without fixing the bankruptcy system, we could face a situation where bankruptcies spiral out of control even beyond what they were in the good times of 1998, 1997, and 1996.").

[91] *See, e.g.*, 147 CONG. REC. H133 (daily ed. Jan. 31, 2001) (statement of Rep. Gekas (R-Pa.)) ("[House Bill 333] is tailored to make certain that anyone who is so overwhelmed by debt, so swamped by the inability to pay one's [sic] obligations . . . [is entitled] to be discharged in bankruptcy, to be free from the debts that so overwhelmed him."); Jones & Zywicki, *supra* note 64, at 207 ("Apocalyptic rhetoric to the contrary, the reality of means-testing is that it will apply only to bankruptcy filers with above median incomes, sufficient disposable income to fund a plan, the ability to repay a substantial portion of their unsecured debt, and no other overriding hardship.").

[92] *See, e.g.*, 147 CONG. REC. S1807–08 (daily ed. Mar. 5, 2001) (statement of Sen. Grassley (R-Iowa)) ("[C]laims that this bill is unbalanced for the creditor and against the

The Bankruptcy Reform Act's likely effects are no less contested than the explanations for current bankruptcy trends. Reformers claim that the Bankruptcy Reform Act will be beneficial to consumers because toughened debt collection will lead to lower borrowing costs. There is some doubt, however, as to whether credit card interest rates are sufficiently competitive to respond to bankruptcy reforms. Despite deep market penetration and intense competition for new customers,[93] credit card issuers generally compete by offering perks such as purchasing discounts and frequent flier miles, not by offering lower interest rates.[94] Thus, while bank borrowing rates fell from 13.4% to 3.5% between 1980 and 1992, the average credit card interest rate rose from 17.3% to 17.8%.[95] In an influential article, economist Lawrence Ausubel has speculated that this "stickiness" occurs because many consumers deceive themselves about the likelihood they will carry a balance on their credit cards: they fail to seek out lower rates though it would be in their interest to do so.[96] Other commentators have suggested that interest rates remain high because the consumers who carry credit card balances are not savvy enough to appreciate the importance of the interest rate, while more sophisticated cardholders generally use cards only for convenience and do not carry account balances.[97]

The disclosure requirements included in the Bankruptcy Reform Act could mitigate such problems by exposing credit cards' use of low minimum payments and high interest rates to keep consumers in debt.[98] It is doubtful, however, whether the inclusion of further fine-print disclosures on credit card materials would have any significant impact on consumer behavior.[99] Moreover, if Ausubel is correct about borrower self-deception, then it is likely that consumers would persist in ignoring interest rate information, permitting card issuers to continue to charge uncompetitive rates.

On the other hand, some reform proponents have argued that the low transaction costs of credit card borrowing justify the high interest rates.[100] There is also evidence that consumers have begun to take note of interest rates, shifting balances between cards to exploit introductory "teaser" rates and reduce monthly payments.[101] If credit card interest rates are

---

debtor are wrong. There are enhanced consumer protection and information and education provisions . . . .").

[93] *See supra* text accompanying note 76.

[94] *See* SULLIVAN ET AL., *supra* note 43, at 254.

[95] *Id.* at 255 (quoting JAMES MEDOFF & ANDREW HARLESS, THE INDEBTED SOCIETY: ANATOMY OF AN ONGOING DISASTER 12 (1996)).

[96] Lawrence Ausubel, *The Failure of Competition in the Credit Card Market*, AM. ECON. REV. Vol. 81, No. 1, at 50, 70–71 (Mar. 1991).

[97] *See* SULLIVAN ET AL., *supra* note 43, at 255.

[98] *See id.* at 247–48.

[99] *See* Ball & Stuart, *supra* note 37, at 1503–04.

[100] *See, e.g.*, Jones & Zywicki, *supra* note 64, at 228–31.

[101] *See* Jeff Bailey & Scott Kilman, *Taking Credit: Here's What's Driving Some Lend-*

economically reasonable, or are becoming so, then responsible borrowers might reap benefits from toughened bankruptcy laws. Still, the remarkable profitability of consumer lending[102] and the prevalence of high interest rates have led some members of Congress to doubt whether "the credit card industry likely would pass on any of the 'savings' from bankruptcy law changes to individual consumers."[103]

There is also debate as to whether the Bankruptcy Reform Act would produce significant benefits for creditors. As Professor Elizabeth Warren has observed:

> Bankruptcy is the ultimate zero-sum system. Creditors compete for the limited dollars of the people who have declared themselves bankrupt. More to one creditor is necessarily less for another.[104]

Yet the Bankruptcy Reform Act avoids choosing between creditors; to the contrary, it flattens distinctions between them by toughening presumptions against dischargeability and blocking the bifurcation of undersecured debts. The result, in the words of Senator Feingold (D-Wis.), "is a bill at war with itself"—more creditors competing for the same funds.[105] This competition could mean that creditors end up collecting little more than they would under current law. One analyst has concluded that increasing secured creditors' claims under Chapter Thirteen by preventing strip-down would "reduce distributions to unsecured creditors in Chapter 13, eliminating them in many cases."[106] This possibility raises further doubts about the effect on interest rates: if unsecured lenders do not benefit from the bill, there will be no cost savings for them to pass on to consumers. On the other hand, if unsecured creditors receive more, secured creditors might receive less than they would have through Chapter Seven reaffirmation agreements, with the result that the costs of secured borrowing—for instance, automobile financing—could rise.

Critics are particularly concerned about the bill's implications for priority creditors such as former spouses and local tax authorities.[107] The

---

ers Crazy: Borrowers Who Think Working Class Is Getting Hip to Lower Interest Rates; Card Surfing Is the Rage, WALL ST. J., Feb. 20, 1998, at A1.

[102] *See* Lawrence Ausubel, *Credit Card Defaults, Credit Card Profits, and Bankruptcy*, 71 AM. BANKR. L.J. 249, 258–59 (1997). In part because of the high interest rates charged, financial institutions earn higher profits from credit cards than any other lending activity. *See id.*

[103] JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 458.

[104] *Bankruptcy Reform Hearing*, *supra* note 47 (statement of Elizabeth Warren).

[105] 145 CONG. REC. S14,067 (daily ed. Nov. 5, 1999) (statement of Sen. Feingold (D-Wis.)) (commenting on an earlier version of the bill in the 106th Congress).

[106] Norberg, *supra* note 34, at 462.

[107] *See, e.g.*, 145 CONG. REC. S14,067 (daily ed. Nov. 5, 1999) (statement of Sen. Feingold (D-Wis.)) ("In too many cases, I am afraid, [the bill] will hinder families' ability to meet other obligations, particularly their obligations to their own children and to local

House Judiciary Committee Report stresses that "[d]omestic support claimants receive a broad spectrum of special protections" under the Bankruptcy Reform Act.[108] Like current law, the bill assigns priority status to family support obligations[109] and taxes,[110] rendering them non-dischargeable along with certain other debts such as educational loans.[111] Critics worry, however, that since the bill expands the definition of non-dischargeable cash advances and consumer loans, it may place "the single mother seeking money for food into direct competition with credit card debt."[112] Furthermore, by pushing debtors into Chapter Thirteen while restricting their right to re-file if their plans fail, the Bankruptcy Reform Act may make it more difficult for priority creditors to collect what they are owed. Roughly two-thirds of Chapter Thirteen plans fail under current law;[113] analysts predict the success rate would remain low under the reform bill.[114] When a debtor's Chapter Thirteen plan fails, creditors may resume efforts to collect outstanding debts, deploying tactics ranging from letters, phone calls, and visits by collection agents to state law collection procedures, such as wage garnishment and asset foreclosure.[115] In such circumstances, families and other priority creditors may lose out in the competition with more assertive and experienced debt collection professionals.[116] Some debtors who might meet their priority obligations following a Chapter Seven discharge may instead end up paying their limited funds to non-priority creditors who are more assertive following the failure of a Chapter Thirteen plan.

A final important argument concerning the likely effects of the Bankruptcy Reform Act pertains to the costs of the bill for litigants and for the bankruptcy system. Since Chapter Thirteen plans require ongoing court supervision, Chapter Thirteen cases generally involve higher lawyer fees and court costs than Chapter Seven cases.[117] Thus, an increase in

---

taxing authorities.").

[108] HOUSE JUDICIARY COMMITTEE REPORT, *supra* note 6, at 11.

[109] *See* H.R. 333, §§ 211–15.

[110] *See id.* § 705.

[111] *See id.* § 220.

[112] JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 475.

[113] *See* Norberg, *supra* note 34, at 439; *see also* SULLIVAN ET AL., *supra* note 43, at 14.

[114] *See* Norberg, *supra* note 34, at 450–51. While the Bankruptcy Reform Act's means-test might be expected to improve the success rate by pushing higher-income debtors into Chapter Thirteen, Norberg's empirical examination of Chapter Thirteen cases found no correlation between income and plan completion. *See id.* Norberg goes on to say conclude that "[t]he essential unpredictability of success in chapter 13 undermines the case for means-testing." *See id.*; *see also* Jean Braucher, *Increasing Uniformity in Consumer Bankruptcy: Means Testing as a Distraction and the National Bankruptcy Review Commission's Proposals as a Starting Point*, 6 AM. BANKR. INST. L. REV. 1, 11 (1998) ("The impact of stricter means testing would likely include a higher failure rate [for Chapter Thirteen cases].").

[115] *See* White, *supra* note 2, at 211–12 (describing creditors' legal remedies against debtors who default).

[116] *See* JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 472–75.

[117] *See* Norberg, *supra* note 34, at 436.

Chapter Thirteen filings may impose higher costs on the bankruptcy system, while also increasing the proportion of debtors' assets that go to lawyers rather than creditors. Additionally, the Bankruptcy Reform Act's means test may be expensive to administer. One study based on an earlier version of the Bankruptcy Reform Act[118] found that only 3.6% of Chapter Seven debtors would pass the means test and qualify as "can-pays" who should instead have been in Chapter Thirteen.[119] If the Bankruptcy Reform Act's means test would filter out only a small proportion of Chapter Seven debtors, then, in the view of some critics, the costs of administering the test may outweigh the benefits.[120]

Whether the means test is worthwhile despite its costs may depend on a value judgment. In the Committee Report on the Bankruptcy Reform Act, both the majority and the dissenters cite the 3.6% figure to support their views; this suggests the two sides assign different values to the social importance of pushing a small proportion of "can-pays" out of Chapter Seven.[121] At a time when many households are taking on unmanageable levels of debt, it may be important for the bankruptcy system to send a strong signal that debtors who can pay must do so.[122] Furthermore, one scholar has suggested that many courts already apply an *ad hoc* income test in making determinations of Chapter Seven abuse. If that is the case, then the universal application of a "mechanical test" could enhance the bankruptcy system's clarity and efficiency.[123]

The administrative hurdle of a means test could possibly lead debtors outside of the 3.6% to file under Chapter Thirteen even though they could qualify for Chapter Seven. As one commentator predicts, "[s]ome

---

[118] H.R. 3150, 105th Cong. (1998).

[119] Marianne B. Culhane & Michaela M. White, *Taking the New Consumer Bankruptcy Model for a Test Drive: Means-Testing Real Chapter 7 Debtors*, 7 AM. BANKR. INST. L. REV. 27, 31 (1999). While research funded by the credit card industry has suggested that more than ten percent of Chapter Seven debtors would qualify as "can-pays," independent academics, as well as the General Accounting Office have criticized these results. *See id.* at 28–31 (describing industry-funded research and the debate over the results). *But see* Jones & Zywicki, *supra* note 64, at 186–200 (criticizing the Culhane & White study and arguing for the industry-funded research).

[120] *See, e.g.*, *Bankruptcy Reform Hearing*, *supra* note 47, at 1 (statement of Prof. Warren) ("A good example of the failure of cost-benefit analysis is evident in the proposed means test.").

[121] *See* HOUSE JUDICIARY COMMITTEE REPORT, *supra* note 6, at 5–6 n.18 (quoting the Culhane & White study to support the proposition that "debtors can, in fact, repay a significant portion of their debts"); *see also* JUDICIARY COMMITTEE DISSENTING VIEWS, *supra* note 9, at 458 (referring to the figure as "a mere 3.6%").

[122] *See* Jones & Zywicki, *supra* note 64, at 191–92 ("[T]o the extent that there are doubts about the administrative savings that would result from a bright-line statutory means-testing requirement or about the number of individuals who would qualify under means-testing, this moral message [that debts must be repaid] must also be put on the scale in favor of means-testing.").

[123] Felsenfeld, *supra* note 13, at 1402; *see also* Jones & Zywicki, *supra* note 64, at 205 ("[S]tatutory-based means testing would substitute a bright-line rule for the current murky standard. In general, bright-line rules tend to *reduce* administrative costs relative to standards and increase the predictability of their application.") (footnotes omitted).

lawyers would avoid the expense of dealing with abuse or eligibility challenges by putting nearly all clients in chapter 13."[124] Lawyers working for low-income clients on a fixed fee might have little incentive to complete the work required to establish Chapter Thirteen eligibility; meanwhile, debtors filing without representation might find the paperwork discouraging or impossible. As a result, debtors who could qualify for Chapter Seven may end up filing under Chapter Thirteen instead. Given the poor success rate of Chapter Thirteen plans, such debtors might spend years cycling in and out of repayment plans, struggling to pay debts that should have been discharged under Chapter Seven.[125]

This possibility indicates what may be at stake in Congress' efforts to overhaul the American bankruptcy system. The United States is unique among modern western states in offering its citizens the right to clear their debts and start over.[126] This "fresh start" policy has traditionally been seen as a component of America's entrepreneurial, free-market culture: bankruptcy discharge promotes risk-taking by protecting failures; it ensures that no individual's productivity is dampened by the prospect of earning income only to pay off old debts.[127] Western European countries have generally opted instead to compensate for economic risks by providing a generous social safety net.[128] A bankruptcy reform that pushes more debtors into unsuccessful Chapter Thirteen plans would place the United States in the novel position of offering its citizens neither the social protections of European states, nor the power to borrow during emergencies and clear debts through bankruptcy.[129]

---

[124] Braucher, *supra* note 114, at 11.

[125] *See id; see also* SULLIVAN ET AL., *supra* note 43, at 253–54 (describing proposals to push debtors out of Chapter Seven as "a social experiment that could get very ugly"). The possibility that many Chapter Seven debtors may end up filing unsuccessful Chapter Thirteen plans may explain why the credit card industry has lobbied so aggressively for a means test. *See* Warren, *supra* note 33, at 503–06 (indicating that the credit card industry advocated means-testing even before specific means test proposals were developed). Some research suggests unsecured creditors would get little more under Chapter Thirteen plans than they do now in Chapter Seven cases. *See supra* text accompanying note 106. Since most Chapter Thirteen plans fail, however, the displacement of Chapter Seven filers into Chapter Thirteen would mean in practice that credit card lenders could at least pursue debtors out of court, where their experience and sophistication may give them an advantage over other creditors. *See supra* text accompanying notes 115–116; *see also* Jones & Zywicki, *supra* note 64, at 191 ("If there really were no benefit to unsecured creditors from channeling more debtors into Chapter 13 payment plans, the credit industry would not be advocating means-testing.") (footnotes omitted).

[126] *See* SULLIVAN ET AL., *supra* note 43, at 257–59.

[127] *See id.* at 258; *see also* Thomas H. Jackson, *The Fresh-Start Policy in Bankruptcy Law*, 98 HARV. L. REV. 1393 (1985) (arguing that the fresh start policy is justified by psychological tendencies to underestimate the social costs of risky decisions and to trade future costs for present benefits); John M. Czarnetsky, *The Individual and Failure: A Theory of Bankruptcy Discharge*, 32 ARIZ. ST. L.J. 393 (2000) (arguing that debt discharge by the bankruptcy system protects entrepreneurship, creating net benefits for society).

[128] *See* SULLIVAN ET AL., *supra* note 43, at 257.

[129] *Cf. id.* at 257–59 (comparing the European and American models and concluding "the need for a more protective consumer bankruptcy law is directly proportional to the

Given that debtor abuse may not be the sole explanation for high bankruptcy levels, and that interest rate reductions may not be the principle effect of restrictions on debt discharge, the Bankruptcy Reform Act shortchanges the complexity of the issues facing the American bankruptcy system. Toughened bankruptcy laws may deter some borrowing, but for any debtors who continue to accumulate unmanageable debts, the bankruptcy system will face the same choice it has faced throughout American history: it can demand that debtors repay what they can, or it can permit them to clear some debts and make a fresh start. Credit card companies and other lenders have earned healthy profits in a system that reflects the latter choice. Indeed, it is unclear creditors could earn much more in a system that restricts debt discharge; after all, debtors' resources are limited, however large their debts.[130] The question the Bankruptcy Reform Act raises is whether such marginal benefits to lenders are worth the cost of preventing millions of families from putting their financial misfortunes behind them.

—*Zachary Price*

---

size of the social safety net and the availability of consumer credit.").

[130] *See id.* at 255 ("[M]uch of the debt discharged in bankruptcy had already been written off by the issuer as 'uncollectible' because the issuer had determined that the debtor was so unlikely ever to have the resources to repay that spending another dollar on debt collection was throwing good money after bad.").

MENTAL HEALTH PARITY

On March 15, 2001 Senators Pete Domenici (R-N.M.) and Paul Wellstone (D-Minn.) introduced Senate Bill 543, the Mental Health Equitable Treatment Act of 2001 ("MHETA"),[1] a measure that would require group health insurance plans to provide benefits for mental health services equal to those offered for medical and surgical services.[2] The proposal attempts to address inequalities in private health insurance plans that currently limit coverage for mental health services. Senators Domenici and Wellstone introduced a nearly identical bill during the 106th Congress, but it was never reported out of committee.[3] This year, the Senate Committee on Health, Education, Labor and Pensions held hearings on the bill in July and then unanimously approved it on August 1 with several minor changes.[4] On October 30, the Senate approved the MHETA as an amendment to the Fiscal Year 2002 appropriations bill for the Departments of Labor and Health and Human Services;[5] the mental health parity provisions are expected to be a point of contention during the conference committee negotiations on the appropriations bill.[6] The White House has indicated that President Bush wants to see a compromise reached that would provide some level of coverage for mental health services, but would not significantly increase the overall costs of health insurance.[7]

Congressional attention to the problem of insurance coverage for mental health services may be spurred in part by a growing public awareness of the incidence and impact of mental illness in the United States.[8] About one in five Americans, over forty million individuals, suf-

---

[1] Mental Health Equitable Treatment Act of 2001, S. 543, 107th Cong. (2001).

[2] *See* 147 CONG. REC. S2393–96 (daily ed. Mar. 15, 2001). For purposes of this Recent Development, "mental health benefits" refers to benefits offered for the treatment of mental health conditions, while "medical and surgical benefits" or "medical benefits" refers to all other benefits offered under a particular health care plan. *See* S. 543, §§ 2–3.

[3] *See* Mental Health Equitable Treatment Act of 1999, S. 796, 106th Cong. (1999); *see also* Library of Congress, *Bill Summary and Status for the 106th Cong., S. 796, at* http://thomas.loc.gov (last visited Nov. 15, 2001).

[4] *See* S. REP. NO. 107-61, at 5–6, 14–15 (2001). Most of the changes were technical in nature. *See id.* at 14–15; *see also infra* notes 78, 81.

[5] *See* 147 CONG. REC. S11,183, 11,231–32 (daily ed. Oct. 30, 2001). The appropriations bill is House of Representatives Bill 3061. H.R. 3061, 107th Cong. (2001).

[6] *See* Janet Hook, *Senate OKs Insurance Measure For Mentally Ill*, L.A. TIMES, Oct. 31, 2001, at A22. The House of Representatives has approved a much more limited version of the MHETA as an amendment to the Economic and Security Recovery Act of 2001. *See* Economic and Security Recovery Act of 2001, H.R. 3090, 107th Cong. (2001); 147 CONG. REC. H7240 (daily ed. Oct. 24, 2001).

[7] *See* Ari Fleischer Press Conference, Transcript #110606CN.V54 (CNN, Nov. 7, 2001).

[8] *Cf.* S. REP. NO. 107-61, at 6; Richard G. Frank, Chris Koyanagi, & Thomas G. McGuire, *The Politics and Economics of Mental Health 'Parity' Laws*, HEALTH AFF., July-Aug. 1997, at 108, 113–14.

fers from a diagnosable mental disorder during the course of a year.[9] This includes one in five adolescents and children.[10] Almost half of adults and children with diagnosable disorders experience significant functional impairment as a result.[11] Yet, approximately two-thirds of individuals with potentially diagnosable disorders do not seek treatment.[12] A majority of both insured and uninsured individuals suffering from untreated mental health disorders cite cost as the primary reason that they do not use mental health services.[13] This may be due in part to unequal health insurance coverage for mental health services, which results in significant cost-shifting from private insurers to individuals.[14]

Untreated mental illnesses are associated with a number of societal problems.[15] It is estimated that mental illnesses cost the United States $113 billion annually in indirect costs.[16] Of this total amount, $8 billion is attributed to crime and welfare costs.[17] Approximately 16% of all inmates in state and federal jails across the nation, over 283,000 individu-

---

[9] *See* REP. NO. 107-61, at 6; *see also* SURGEON GENERAL, U.S. DEP'T OF HEALTH & HUMAN SERVS., MENTAL HEALTH: A REPORT OF THE SURGEON GENERAL 46 (1999), *available at* http://www.surgeongeneral.gov/library/mentalhealth/home.html (last visited Nov. 25, 2001) [hereinafter SURGEON GENERAL].

[10] SURGEON GENERAL, *supra* note 9, at 46–47.

[11] *Id.* Specifically, the Surgeon General's Report estimates that every year 9% of adults suffer from a mental health disorder resulting in significant functional impairment, and that 5 to 9% of children suffer from a mental health disorder that can be characterized as a "serious emotional disturbance." *Id.*

[12] *Id.* at 75.

[13] *Id.* at 409. In a telephone survey by the Robert Wood Johnson Foundation, 55% of insured respondents who perceived their own need for mental health treatment and 83% of uninsured respondents in this category cited cost as the primary reason they do not seek treatment. *Id.*

[14] *See id.* at 418. A recent study compared out-of-pocket costs for privately insured individuals under their current mental health coverage with their potential out-of-pocket costs under parity coverage (defined as equal to current coverage for medical services). *See* Samuel H. Zuvekas, Jessica S. Banthin & Thomas M. Selden, *Mental Health Parity: What Are the Gaps in Coverage?*, 1 J. MENTAL HEALTH POL'Y & ECON. 135, 139 (1998). The study compared these costs under several different scenarios, each assuming a different level of use for mental health services. *See id.* The study found that, based on current coverage, patients on average would have to pay about 30 to 40% of the total costs of mental health services, compared to less than 20% under parity coverage. *See id.* Moreover, the gaps in coverage increased as total expenditures rose, so that a patient incurring $60,000 in mental health service costs under current coverage would have to pay about 44% out-of-pocket, while a patient incurring an equal amount in total costs under parity coverage would have to pay only 3% out-of-pocket. *See id.*

[15] While the focus of the congressional debate over mental health parity has been on the potential for costs savings in the workplace, *see infra* text accompanying notes 90–95, supporters also note that untreated mental illnesses are associated with many other social costs. *See, e.g.*, 147 CONG. REC. S11,166 (daily ed. Oct. 30, 2001) (statement of Sen. Domenici).

[16] *See* NATIONAL MENTAL HEALTH ASSOCIATION, LABOR DAY 2001 REPORT: UNTREATED AND MISTREATED MENTAL ILLNESS AND SUBSTANCE ABUSE COSTS THE U.S. $113 BILLION A YEAR; THE 'DOLLARS AND SENSE' CASE FOR INCREASED INVESTMENTS IN MENTAL HEALTH AND SUBSTANCE ABUSE 3, *available at* http://www.nmha.org/pdfdocs/laborday2001.pdf.

[17] *See id.*

als, are suffering from a serious mental illness, such as schizophrenia, bipolar disorder, or major depression.[18] In addition, the mentally ill experience higher rates of unemployment than the general population.[19] Approximately one-fourth of unemployed Americans receiving federal disability benefits, two million individuals, are considered disabled because of mental illnesses.[20] Even more striking, 40% of the homeless in the United States are adults with severe mental illnesses.[21] Finally, the effects of mental illness may serve to exacerbate existing racial inequalities. A report released by the Surgeon General this year found striking racial disparities in access to mental health services: minority individuals have poorer access to mental health services, are less likely to receive mental health services, often receive lower quality care when they do seek treatment, and are underrepresented in mental health research studies.[22] As a result of unequal treatment opportunities, "racial and ethnic minorities bear a greater burden from unmet mental health needs and thus suffer a greater loss to their overall health and productivity."[23]

Despite these social costs, very few employers offer mental health benefits at levels equal to medical and surgical benefits in their group health plans.[24] Inequalities in coverage fall within three basic categories: (1) lower limits on the number of inpatient days or outpatient visits per year, (2) lower annual and/or lifetime caps on total expenditures covered, and (3) higher financial requirements such as deductibles, co-payments or co-insurance.[25] A typical group health plan limits coverage for mental health services to thirty hospital days and twenty to twenty-five outpatient visits annually, compared to unlimited coverage for medical and surgical services.[26] Beneficiaries generally are limited to $10,000 to

---

[18] *See* National Alliance for the Mentally Ill, *Where We Stand: The Criminalization of People With Mental Illness* (2001), *at* http://www.nami.org/update/unitedcriminal.html.

[19] *See* NATIONAL MENTAL HEALTH ASSOCIATION, *supra* note 16, at 4 (reporting results of a January 2001 survey finding that the unemployment rate among adults with depression is nearly four times the average rate across the general population).

[20] *See* National Alliance for the Mentally Ill, *Where We Stand: Employment, Work and Income Supports for People With Brain Disorders* (2001), *at* http://www.nami.org/update/unitedemploym.html.

[21] *See* National Alliance for the Mentally Ill, *Where We Stand: Housing* (2001), http://www.nami.org/update/unitedhousing.html.

[22] *See generally* SURGEON GENERAL, U.S. DEP'T OF HEALTH & HUMAN SERVS., MENTAL HEALTH: CULTURE, RACE, AND ETHNICITY, A SUPPLEMENT TO MENTAL HEALTH (2001), *available at* http://www.surgeongeneral.gov/library/mentalhealth/cre/sma-01-3613.pdf.

[23] *See id.* at 17.

[24] *See, e.g.*, Jeffrey A. Buck, et al., *Behavioral Health Benefits in Employer-Sponsored Health Plans*, HEALTH AFF., Mar.-Apr. 1999, at 67, 68–70; M. Audrey Burnam & Jose J. Escarce, *Equity in Managed Care for Mental Disorders*, HEALTH AFF., Sept.-Oct. 1999, at 22, 25.

[25] *See, e.g.*, GENERAL ACCOUNTING OFFICE, MENTAL HEALTH PARITY ACT: DESPITE NEW FEDERAL STANDARDS, MENTAL HEALTH BENEFITS REMAIN LIMITED 12–13 (2000) [hereinafter GENERAL ACCOUNTING OFFICE]; Buck et al., *supra* note 24, at 69–72.

[26] *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 6–7; *see also* Burnam & Escarce, *supra* note 24, at 25; Daniel P. Gitterman, Roland Sturm & Richard M. Scheffler,

$20,000 annually and $25,000 to $60,000 in lifetime mental health benefits, compared to limits of at least $1 million for medical benefits.[27] Patients receiving mental health services also typically pay a larger percentage of the total costs of care due to higher deductibles and co-payments, and the gap between mental health coverage and coverage for medical services widens as total expenditures rise.[28]

These disparities in coverage have worsened in recent years. In 1981, 58% of employers provided coverage for inpatient mental health services equal to coverage for medical services and 10% provided equal outpatient services; by 1993, however, those numbers had fallen significantly to 16% and 4%, respectively.[29] The policy goal of ending these differences in coverage and requiring insurers to offer health benefits on equal terms for mental health services is referred to generally as "mental health parity" or "parity."[30]

Congress first responded to inequalities in mental health insurance coverage in 1996 with the Mental Health Parity Act ("MHPA"),[31] a limited measure that required group health plans to equalize aggregate lifetime and annual dollar limits for mental health benefits with comparable limits imposed on medical and surgical benefits.[32] The scope of the MHPA was restricted in many respects. First, the law made it clear that employers were not required to offer mental health benefits at all.[33] If such benefits were provided, the MHPA allowed "mental health benefits"

Toward Full Mental Health Parity and Beyond, HEALTH AFF., July-Aug. 2001, at 68, 69.

[27] *See* Buck et al., *supra* note 24, at 72; *see also* Burnam & Escarce, *supra* note 24, at 25; Roland Sturm & Joyce McCulloch, *Mental Health and Substance Abuse Benefits in Carve-Out Plans and the Mental Health Parity Act of 1996*, J. HEALTH CARE FIN., Mar. 22, 1998, at 82, 86–88.

[28] *See* Sturm & McCulloch, *supra* note 27, at 88–92; *see also* Zuvekas, Banthin & Selden, *supra* note 14. The gap in coverage increases with higher expenditures due to differences in lifetime or annual limits, as well as the fact that medical benefits are typically covered by a "stop-loss" provision that places a cap on total out-of-pocket costs for patients (after which the health plan picks up 100% of additional costs), while mental health benefits generally are not protected by such a cap. *Id.*

[29] ALAN L. OTTEN, MILBANK MEMORIAL FUND, MENTAL HEALTH PARITY: WHAT CAN IT ACCOMPLISH IN A MARKET DOMINATED BY MANAGED CARE? 3 (1998), *available at* http://www.milbank.org/ mrparity.html (citing figures from the Bureau of Labor Statistics). This increasing inequality may, in part, be a response to trends in health care spending. *See id.* at 2–3. During the 1970s and 1980s, insurance costs for mental health services rose at up to two times the annual rate for overall health care costs. *Id.* at 2. More recent data indicate that the gaps in coverage continue to widen. *See* Buck et al., *supra* note 24, at 68.

[30] *See* SURGEON GENERAL, *supra* note 9, at 547–58.

[31] *See* Mental Health Parity Act of 1996, Pub. L. No. 104-204, § 701-03, 110 Stat. 2944–50 (1996) [hereinafter Mental Health Parity Act] (to be codified at 29 U.S.C. § 1185a and 42 U.S.C. § 300gg-5).

[32] *See id.* The law offered health plans several options for eliminating existing differences in limits between medical and mental health plans. Plans could either provide no annual or lifetime limits on either type of benefits, apply a single lifetime and annual limit to all benefits, or increase the annual and lifetime limits on mental health benefits so that they were equal to comparable limits imposed on medical and surgical benefits. *See id.* § 702–03.

[33] *See id.*

to be defined differently under the terms and conditions of individual plans; plans could exclude certain categories of mental illness from coverage.[34] Benefits for substance abuse and chemical dependency were explicitly exempt.[35] In addition, the law did not apply to Medicare beneficiaries and those covered by individual insurance plans or plans offered by employers with fewer than fifty employees.[36] Businesses also could apply for an exemption from the provisions of the MHPA if they could demonstrate that compliance with the law would increase their overall group health plan costs by more than 1%.[37] Finally, and most significantly, the 1996 Act did not reach many of the tools used by health insurers to limit mental health coverage, such as higher co-payments and deductibles and lower limits on inpatient days and outpatient visits.[38] The 1996 law became effective on January 1, 1998 and contained a sunset provision that expired on September 30, 2001.[39]

The results of a recent study by the General Accounting Office ("GAO") indicate that, while the vast majority of employers have complied with the MHPA, they have also responded by tightening coverage of mental health benefits in other areas, relying on the types of restrictions left untouched by the law.[40] Eighty-six percent of employers surveyed in 1999 reported that their group health plans were in compliance with the federal law.[41] For most of these employers, compliance with the MHPA had no noticeable effect on their claims' costs.[42] This may be due in part, however, to employers' increased reliance on the types of benefits restrictions still available under the MHPA, such as higher co-payments

---

[34] *Id.*

[35] *See id.*

[36] *See id.*

[37] *Id.* The implementation of this provision was contested, with some arguing unsuccessfully that businesses should be able to apply for exemptions based on projected cost estimates rather than data collected from actual experience under the law. *See, e.g.*, Carolyn M. Levinson & Benjamin G. Druss, *The Evolution of Mental Health Parity in American Politics*, ADMIN. & POL'Y IN MENTAL HEALTH, Nov. 2000, at 139, 142. Interim regulations established that employers would have to comply with the law for six months and then submit data from their actual experience operating under the law in order to qualify for an exemption. *See* Interim Rules for Mental Health Parity, 62 Fed. Reg. 66,932, 66,934–37 (Dec. 22, 1997) (to be codified at 29 C.F.R. 2590, 45 C.F.R. 146, and 26 C.F.R. 54). In the end, only twelve private companies and five non-federal public companies applied for financial waivers. S. REP. No. 106-71, at 3 n.3.

[38] *See* Mental Health Parity Act §§ 702–703.

[39] *Id.* § 703.

[40] *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 4. The GAO study surveyed 1656 employers in the states that had less comprehensive laws than the MHPA at the time of its passage in order to examine the effects of the law on insurance coverage. *See id.*

[41] *Id.* at 10–11. By comparison, before the Mental Health Parity Act, only 55% of employers had provided equal annual and lifetime dollar limits for mental health coverage. *Id.*

[42] *See id.* at 16–17. Sixty percent of employers said that they did not know whether compliance increased costs, 37% reported that compliance did not raise costs, and only 3% indicated that compliance had increased costs. *Id.* at 16. Less than 1% of employers reported that they had dropped their mental health or general health insurance benefits because of the new law. *Id.* at 17.

and tighter limits on inpatient care days.[43] Eighty-seven percent of plans in compliance contained other limits on mental health benefits, and there was evidence that many plans had recently tightened these restrictions.[44] Sixty-five percent of employers who had to expand their mental health coverage in order to comply with the MHPA also tightened another plan feature during the same time period.[45] By comparison, only 26% of employers who either were already in compliance in 1996 or still have not complied had further restricted their coverage of mental health benefits during this period.[46]

Several other measures since 1996 have expanded the scope of federal mental health parity requirements. Provisions in the 1997 Balanced Budget Act[47] require Medicaid managed care plans and state Children's Health Insurance Programs[48] to comply with the requirements of the MHPA.[49] Even more significantly, on June 7, 1999 at a White House conference on mental health, then-President Clinton issued an executive order to the Office of Personnel Management requiring the implementation of full parity[50] for both mental health and substance abuse services in the Federal Employees Health Benefits Program ("FEHBP") by January 1, 2001.[51] This federal program provides coverage to over 8.5 million fed-

---

[43] *See id.* at 14–15.

[44] *See id.* at 12. The most common features in compliant plans were fewer outpatient visits allowed (65% of compliant plans) and fewer inpatient days allowed (65%). *Id.* at 12. Other limitations included higher outpatient co-payments (27%) and co-insurance (25%), and higher inpatient co-insurance (10%) and co-payments (5%). *Id.* at 12–13. The study reports that non-compliant plans already contained nearly identical limitations on coverage. *See id.* at 13.

[45] *See id.* at 14–15. The most common changes were to lower limits on outpatient visits (51% of newly compliant employers) and inpatient days (36). *See id.* Other surveys confirm that health plans continue to rely on day or visit limits and higher rates of cost-sharing to limit access to mental health services. *See, e.g.*, Donald P. Gitterman, Roland Sturm, Rosalie Liccardo Pacula & Richard M. Scheffler, *Does the Sunset of Mental Health Parity Really Matter?*, ADMIN. & POL'Y IN MENTAL HEALTH, May 2001, at 353, 362–63; *see also* Gitterman, Sturm & Scheffler, *supra* note 26, at 69.

[46] GENERAL ACCOUNTING OFFICE, *supra* note 25, at 14–15.

[47] Balanced Budget Act of 1997, Publ. L. No. 105-33, 111 Stat. 251.

[48] *See* 42 U.S.C. § 1397aa (1994 & Supp. III 1998). The State Children's Health Insurance Program provides federal funds for states to provide health insurance coverage to uninsured, low-income children. *See id.* To qualify for federal funds, state plans must meet federal coverage mandates, including the mental health parity requirement. *See id.*

[49] *See* 42 U.S.C. § 1396u-2(b)(8) (2000); 42 U.S.C. § 1397cc(f)(2) (2000).

[50] For purposes of this Recent Development, "full parity" refers to any parity mandate requiring that all benefits for mental health services be set equal to benefits for medical and surgical services.

[51] *See* Office of Personnel Mgmt., *Mental Health and Substance Abuse Parity in the Federal Employees Health Benefits Program, at* http://www.opm.gov/insure/health/parity (last visited Nov. 15, 2001). Guidelines issued by the Office of Personnel Management require carriers to offer full parity for in-network mental health services if the services are part of an authorized treatment plan, delivered according to standard protocols, and meet the plan's medical necessity criteria. *See* Letter 2000-17 from the Office of Personnel Mgmt., to all FEHBP Program Carriers (Apr. 11, 2000), *at* http://www.opm.gov/insure/2000-17.pdf. These guidelines encourage insurance carriers participating in the FEHBP to use managed care tools to control costs. *See id.* at 2–3. Prior to former President Clinton's

eral employees, retirees, and their family members.[52] The FEHBP provisions have been cited as a model for extending full parity at the federal level.[53]

Some of the most dramatic legislative changes in recent years, however, have occurred at the state level. When the MHPA became law in 1996 only five states had enacted legislation requiring some form of mental health parity in health insurance coverage for private employees.[54] Another four states required some level of parity for mental health services in the benefits offered to state employees.[55] The pace of reform picked up at the state level after the passage of the MHPA, with over thirty states mandating mental health coverage between 1997 and 2001.[56] The result of this reform movement is that, by August 2001, fewer than a dozen states did not have some form of mental health parity legislation.[57] The MHPA may have encouraged this state legislative activity in at least two ways: by highlighting the issue of mental health parity[58] and by providing an incentive for states to head off federal regulatory oversight by enacting laws that were either comparable to the federal law or more comprehensive.[59]

---

announcement, the FEHBP already mandated a minimum package of mental health benefits for participating plans in 1994 and eliminated disparities in lifetime and annual dollar limits on benefits in 1995. *See Achieving Parity for Mental Health Treatment: Hearing on S. 543 Before the Senate Comm. on Health, Educ., Labor, & Pensions*, 107th Cong. 3–4 (forthcoming 2001) [hereinafter *Achieving Parity for Mental Health Treatment Hearings*] (statement of William E. Flynn, III, Assoc. Dir. for Retirement & Ins., Office of Personnel Mgmt), *available at* http://www.senate.gov/%7Elabor/107hearings/july2001/071101wt/071101wt.htm.

[52] *Achieving Parity for Mental Health Treatment Hearings*, *supra* note 51 at 2 (statement of William E. Flynn, III).

[53] *See, e.g.*, S. REP. NO. 107-61, at 2; *see also* 147 CONG. REC. S2393 (daily ed. Mar. 15, 2001) (statement of Sen. Domenici); 147 CONG. REC. S2396 (daily ed. Mar. 15, 2001) (statement of Sen. Wellstone).

[54] *See* RUTH L. KIRCHSTEIN, NAT'L INSTITUTES OF HEALTH, INSURANCE PARITY FOR MENTAL HEALTH: COST, ACCESS, AND QUALITY, FINAL REPORT TO CONGRESS BY THE NATIONAL ADVISORY MENTAL HEALTH COUNCIL 36–45 (2000). These were the first state laws mandating that benefits for mental health services be set equal to benefits for medical services; some states already had laws requiring insurers to offer a bare minimum of mental health benefits. *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 40–61.

[55] *See* KIRCHSTEIN, *supra* note 54, at 36–45.

[56] *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 40–61. Even these numbers do not capture the level of state activity on mental health parity because, in recent years, many states also have been considering measures that would expand existing laws. In 2001, thirty-four states considered seventy-six separate legislative proposals on mental health parity, and thirteen states passed some form of legislation. *See* NATIONAL ALLIANCE FOR THE MENTALLY ILL, STATE PARITY LEGISLATION 2001: TRACKING REPORT, *at* http://www.nami.org/update/sparity.html (last modified Aug. 1, 2001).

[57] *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 40–61.

[58] *See, e.g.*, Gitterman, Sturm & Scheffler, *supra* note 26, at 68 ("[The MHPA] did place parity on the agenda and enabled state legislatures to experiment with more-comprehensive parity provisions.").

[59] *See* Interim Rules for Mental Health Parity, 62 Fed. Reg. at 66,934–37. Under the MHPA, state insurance regulators are allowed to retain authority for implementing parity over insurance carriers in their state in accordance with state law, as long as their parity laws either meet or exceed the requirements of federal law. *See id.* Over a dozen states

The provisions of state parity laws differ substantially from one jurisdiction to another, with many states mandating broader coverage than the MHPA. More than half of the states with parity laws have requirements that exceed the federal Act, and over twenty states mandate full parity.[60] Nonetheless, the scope of state measures is still limited in a number of ways. Many state laws mandate parity only for specified serious mental illnesses.[61] Almost half of the state laws also contain exemptions for small employers, employers who can demonstrate a specified increase in health plan costs, or both.[62] The greatest limit to the reach of state laws, however, is a provision in the federal Employee Retirement Income Security Act of 1974 ("ERISA"),[63] which has been interpreted to preempt state regulation of self-insured employer health plans.[64] Of the 130 million Americans covered by employer-sponsored health insurance, approximately 43%, or fifty-six million Americans, receive coverage through self-insured plans that are exempt from state regulation.[65]

The call for further congressional action on mental health parity in the 107th Congress is based on at least three considerations. First, the MHPA expired on September 30, 2001, reversing the only comprehensive federal mandate on mental health parity. Second, as sponsors of the MHETA have argued, simply reauthorizing the MHPA would leave intact a variety of restrictions on mental health care that have become even more pervasive in recent years.[66] Finally, although most states have en-

---

initially passed statutes simply meeting the federal requirements. *See* Kevin D. Hennessy & Howard H. Goldman, *Full Parity: Steps Toward Treatment Equity for Mental & Addictive Disorders*, HEALTH AFF., July-Aug. 2001, at 58, 61.

[60] *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 40–61.

[61] *See id.*; *see also* KIRCHSTEIN, *supra* note 54, at 36–45; BAZELON CENTER FOR MENTAL HEALTH LAW, PARITY IN MENTAL HEALTH INSURANCE COVERAGE, *at* http://www.webcom.com/ bazelon/stateswithparity.html (last visited Nov. 17, 2001).

[62] *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 40–61; *see also* KIRCHSTEIN, *supra* note 54, at 36–45; BAZELON CENTER FOR MENTAL HEALTH LAW, *supra* note 61.

[63] Employee Retirement Income Security Act of 1974, 29 U.S.C. § 1144 (1999).

[64] *See Metro. Life Ins. Co. v. Mass.*, 471 U.S. 724 (1985). Self-insured plans are those in which an employer funds its own health insurance plan and assumes all of the financial risk, rather than purchasing a group policy from a third-party insurance provider. *See* S. REP. NO. 107-61, at 3 n.4.

[65] *See* S. REP. NO. 107-61, at 4 n.7. Even this figure may overstate the potential reach of state insurance regulations. A study commissioned by the Health Insurance Association of America estimates that on average only 33 to 42% of a state's population will be affected by state insurance regulations because these laws do not reach the uninsured or those covered by federal public insurance systems, individual insurance plans, or self-insured employer plans. *See* GAIL A. JENSEN & MICHAEL A. MORRISEY, HEALTH INSURANCE ASSOCIATION OF AMERICA, MANDATED BENEFIT LAWS AND EMPLOYER-SPONSORED HEALTH INSURANCE 11 (1999).

[66] *See, e.g.*, *Achieving Parity for Mental Health Treatment Hearings*, *supra* note 51 (statement of Sen. Paul D. Wellstone); *Mental Health Parity: Hearings on S. 796 Before the Senate Comm. on Health, Educ., Labor, & Pensions*, 106th Cong. 3 (May 18, 2000) [hereinafter *Mental Health Parity Hearings*] (statement of Sen. Kennedy); 147 CONG. REC. S2393 (daily ed. Mar. 15, 2001) (statement of Sen. Domenici). Even at the time of the MHPA's passage, Senator Wellstone cited the need for further legislation to implement full parity:

acted some form of parity legislation at this point, the scope of these laws is limited by restrictive definitions of mental illness, exemptions for various employers, and the preemptive effect of ERISA on self-insured plans.[67] Only federal legislation can establish a comprehensive and far-reaching set of standards.

The MHETA would respond to this need by requiring full parity for coverage of mental health services in group health plans, significantly expanding the scope of parity requirements across the nation. Beginning on January 1, 2003,[68] the Act would prohibit group health plans that provide both medical and mental health benefits from imposing any treatment limitations or financial requirements on mental health benefits without imposing comparable limitations on medical and surgical benefits.[69] Health plans would be required to provide equal coverage for all mental health conditions listed in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition ("DSM IV-TR").[70] Employers would still have some leeway in setting benefits, however, by virtue of their continued ability to control the definition of covered services under the general terms and conditions of their plans.[71] The MHETA provisions

---

> [W]hile the passage of this amendment was a historic step forward for people with mental illnesses, the amendment was a first step and a first step only. It does not require parity for copayments or deductibles or inpatient days or outpatient visit limits. It also does not include substance abuse services . . . . We have much more work to do and I look forward to consideration of legislation which would provide full parity coverage for mental health and substance abuse services.

142 CONG. REC. S11,569 (1996) (statement of Sen. Wellstone). *See also* 142 CONG. REC. S9917 (1996) (statement of Sen. Domenici) ("Frankly . . . from where we started, I will confess to everyone, this compromise truly-truly-dramatically reduced our expectations and our hopes.").

[67] *See, e.g., Achieving Parity for Mental Health Treatment Hearings, supra* note 51 (statement of Sen. Paul D. Wellstone).

[68] The Act as originally written would have been effective on January 1, 2002. S. 543, §§ 2–3. Implementation was delayed when the Act was introduced as an amendment to the Department of Health and Human Services appropriations bill, in order to allow more time for compliance and to ensure that parity costs for the federal government would not affect the Fiscal Year 2002 budget. *See* 147 CONG. REC. S 11,173 (daily ed. Oct. 30, 2001).

[69] *See* S. 543, §§ 2–3. Financial requirements are defined to include deductibles, co-insurance, co-payments, other forms of cost-sharing between employers and employees, and annual or lifetime limits on benefits paid. *See id.* Treatment limitations include restrictions on the frequency of treatment, number of inpatient or outpatient visits, or days of coverage. *See id.*

[70] *See id.* DSM-IV-TR is the most recent update of a comprehensive diagnostic and classification system for mental health conditions published by the American Psychiatric Association, and is considered the industry standard for mental health professionals in the United States. *See* American Psychiatric Association, *DSM Future Activities: Diagnostic and Statistical Manual of Mental Disorders, at* http://www.psych.org/clin_res/dsm/index.cfm (last visited Nov. 15, 2001). The DSM-IV-TR definition was chosen because of its broad scope in order to ensure "access to mental health benefits across the full range of diagnoses and eliminate discrimination on the basis of a specific mental illness, disorder or diagnosis." *See* S. REP. No. 107-61, at 16.

[71] *See* S. 543, §§ 2–3.

would extend only to mental health services that meet the criteria of medical necessity as defined and set forth by each group health plan.[72] On the other hand, the Act is explicit in stating that, where existing state laws provide greater protections to employees, these laws would not be preempted by the new federal provisions.[73] Finally, the MHETA calls for a post-implementation GAO study to examine the Act's effects on cost and quality of care and access to care.[74]

Although the MHETA would go a long way toward providing for equal access to mental health services, it still contains significant limitations.[75] Most significantly, health plans would not be required to offer mental health benefits at all, and, even when they did, there would be no guarantee that specific services would be offered as long as any restrictions in coverage applied equally to both medical and mental health benefits.[76] In addition, the MHETA continues the MHPA practice of excluding Medicare beneficiaries, those insured under individual plans,[77] small businesses employing fifty or fewer individuals,[78] and those need-

---

[72] *See id.*

[73] *See id.* § 4.

[74] *See id.* § 5.

[75] A comparison with House Bill 162 is illustrative. The bill, introduced by Representative Marge Roukema (R-N.J.), contains many of the same requirements as the MHETA but reaches further in its scope. *See* Mental Health and Substance Abuse Parity Amendments of 2001, H.R. 162, 107th Cong. (2001). The House measure would extend parity to benefits for substance abuse and chemical dependency, and would require providers in the individual health plan market to comply with the same parity provisions after an additional six months. *See id.* House Bill 162 currently has over 200 co-sponsors. Library of Congress, *Bill Summary & Status for the 107th Congress, H.R. 162, Co-Sponsors*, http://thomas.loc.gov (last visited Nov. 15, 2001). Roukema has disagreed with Senators Domenici and Wellstone in the past about how comprehensive a proposed parity bill should be in order to maximize its chances of passage. *See* Dana Bazelon, *Mental Health Coverage Gains Support on Hill*, THE HILL, June 23, 1999, at 29.

[76] *See* S. 543, §§ 2–3.

[77] *See id.* Under current law, Medicare imposes a 190-day lifetime limit on inpatient care for mental health services without imposing any limits on inpatient medical services, covers only 62.5% of mental health expenses compared to 80 to 100% of medical expenses, and allows for higher co-payments for outpatient mental health care compared to medical care. *See* 42 U.S.C. §§ 1395d, 1395l (1999). Legislation has been introduced to require non-discriminatory coverage under Medicare. Senate Bill 690 and House Bill 1522 would require full parity for mental health services under Medicare. *See* Medicare Mental Health Modernization Act of 2001, S. 690, 107th Cong. (2001); H.R. 1522, 107th Cong. (2001). Senate Bill 841 and House Bill 599 would provide more limited parity by barring discriminatory co-payments for outpatient care under Medicare. *See* Medicare Mental Illness Nondiscrimination Act of 2001, S. 841, 107th Cong. (2001); H.R. 599, 107th Cong. (2001).

[78] *See* S. 543, §§ 2–3. The original bill defined small businesses as those employing twenty-five or fewer workers, but the number was raised during the committee mark-up. *See* S. REP. 107-61, at 14–15, 18–19. This seemingly minor change is actually quite significant; the National Alliance for the Mentally Ill estimates that the change will exempt plans covering an additional fifteen million individuals. National Alliance for the Mentally Ill, *Parity in Insurance Coverage: Where We Stand*, *at* http://www.nami.org/update/unitedparity.html (last visited Nov. 15, 2001).

ing treatment for substance abuse and chemical dependency.[79] Another shortfall of the Act is that health plans may still limit benefits by using managed care techniques—such as utilization review and criteria for medical necessity and appropriateness of care—and there is no guarantee that these tools will be applied equally to mental health and medical benefits.[80] Finally, the scope of the MHETA is limited by the fact that plans may continue to apply unequal restrictions to mental health benefits for out-of-network services, as long as their in-network services comply with the Act.[81]

Advocates frame mental health parity as a call to end unwarranted discrimination against the mentally ill.[82] After several decades of research and advances in scientific understanding of mental illness, it is now clear that most mental health disorders have a biological basis and can be treated with a range of remedies that are just as effective as those for physical disorders.[83] Proponents of parity argue that, particularly given these advances in research, there is no scientific justification for treating mental health services differently than general medical services.[84] In his statement introducing the MHETA, Senator Domenici focused on this theme:

> [S]adly, those suffering from a mental illness do not enjoy those same benefits of treatment and medical advances because all too often insurance discriminates against illnesses of the brain. Individuals suffering from a mental illness face this discrimination even though medical science is in an era where we can accu-

---

[79] *See* S. 543, §§ 2–3. The limitation on substance abuse and chemical dependency treatment is significant because 3% of Americans (15% of those with mental health conditions) suffer from both mental health disorders and substance abuse or chemical dependence. *See* SURGEON GENERAL, *supra* note 9, at 46. These individuals with co-occurring disorders are more likely to experience chronic symptoms and to utilize treatment services. *See id.*

[80] *See* S. 543, §§ 2–3.

[81] *See id.* This provision was one of the technical changes added by the committee mark-up. *See* S. REP. NO. 107-61, at 14–15, 18–19. The bill specifies that health plans must provide "reasonable access" to in-network providers and facilities. *See* S. 543, §§ 2–3. According to the Senate report on the MHETA, this means "comparable access as provided for medical and surgical benefits." S. REP. NO. 107-61, at 16. Plans that significantly increase restrictions on out-of-network services will "be viewed as violating the spirit of the law." *Id.*

[82] *See, e.g.*, Sen. Edward M. Kennedy, *Mental Health Services is a Civil Rights Issue*, THE HILL, July 18, 2001, at 44 (describing access to mental health services as "one of the most important civil rights issues facing the nation.").

[83] *See* SURGEON GENERAL, *supra* note 9, at 5–6, 9–11. The body of the Surgeon General's report provides an overview of recent research regarding the brain generally and specific mental disorders. *See id.* at 13–23.

[84] *See, e.g.*, *Achieving Parity for Mental Health Treatment Hearings, supra* note 51 (statement submitted by Jacqueline Shannon, President, National Alliance for the Mentally Ill), *available at* http://www.nami.org/update/20010730.html (last visited Jan. 1, 2002); 147 CONG. REC. S2393–94 (daily ed. Mar. 15, 2001) (statement of Sen. Domenici).

> rately diagnose mental illnesses and treat those afflicted so they can be productive.[85]

Supporters of parity contend that current limits on mental health coverage in insurance plans are inefficient and arbitrary tools for rationing health care, creating an unjustifiable obstacle to medically necessary and appropriate care.[86]

Supporters of mental health parity stress two other discriminatory effects of the current gaps in mental health coverage. First, they contend that strict limits on insurance coverage for mental health services contribute to the sense of discrimination and stigma facing mentally ill individuals; this ostracizing effect in itself becomes a barrier to treatment.[87] They also claim that the inequitable treatment of mental health needs spreads to the families of the mentally ill, who have to assume significant out-of-pocket costs because they lack adequate health insurance coverage.[88]

The second major argument put forward by advocates of mental health parity is that full parity can reduce some of the high costs that untreated mental illnesses impose on society, ultimately benefiting both employers and consumers. In addition to the societal costs discussed above,[89] mental illnesses result in significant expenses for employers through workplace disability.[90] According to a comprehensive study by

---

[85] 147 CONG. REC. S2393. (daily ed. Mar. 15, 2001) (statement of Sen. Domenici); *see also* 147 CONG. REC. S2395–96 (daily ed. Mar. 15, 2001) (statement of Sen. Wellstone) ("The discrimination continues despite the fact that there is no biomedical justification for differentiating serious mental illness from other serious and potentially chronic disorders, nor for judging mental disorders to be in any way less real or less deserving of treatment."). Former Speaker of the House Newt Gingrich (R-Ga.), who opposed parity during the 1990s, similarly commented in a 1999 editorial that new government-funded research on the human brain "will prove that mental health parity is essential to any health policy and will offer opportunities to cure schizophrenia, bipolar disease, Alzheimer's and other current challenges." Newt Gingrich, *We Must Fund the Scientific Revolution*, WASH. POST, Oct. 18, 1999, at A19.

[86] *See* Burnam & Escarce, *supra* note 24, at 25. The sheer variety of limitations used is also inefficient from an administrative point of view. A nationwide survey of health plans by Roland Sturm and Joyce McCulloch found at least 1,229 different types of restrictions across 4,160 different plans, suggesting that these restrictions are haphazard and not strategically tailored to achieve their goals. *See* Sturm & McCulloch, *supra* note 27, at 85–86.

[87] *See, e.g.*, *Achieving Parity in Mental Health Services, supra* note 51 (statement of Stanford J. Alexander, Chairman, Weingarten Realty Investors); SURGEON GENERAL, *supra* note 9, at 8; Hennessy & Goldman, *supra* note 59, at 65 ("Unfortunately, however, such external obstacles can be easily embraced by these individuals as further evidence of society's belief that their illness is somehow less real, less debilitating, and less worthy of treatment than is the illness of someone with a physical health condition.").

[88] *See, e.g.*, *Achieving Parity for Mental Health Treatment Hearings, supra* note 51 (statement submitted by Jacqueline Shannon) *available at* http://www.nami.org/update/20010730.html.

[89] *See supra* text accompanying notes 15–23.

[90] *See, e.g.*, NATIONAL MENTAL HEALTH ASSOCIATION, *supra* note 16, at 3; SURGEON GENERAL, *supra* note 9, at 411–12.

the World Health Organization, four of the ten leading causes of disability worldwide are mental disorders, and all mental disorders combined account for approximately 15.4% of the overall burden of disease in established market economies worldwide.[91] Of the estimated $113 billion annually in indirect costs from mental illnesses in the United States, over 90% is due to lost productivity.[92] Another study finds that mental illnesses result in about one billion days of lost productivity annually in the United States.[93] Studies and testimonial evidence from employers that have already experimented with full mental health parity indicate that greater access to mental health services can result in lower disability claims, higher productivity, and lower rates of absenteeism.[94] In addition, increased use of mental health services in insurance groups correlates with a comparable reduction in overall insurance expenditures on general health services.[95]

While parity would represent an important symbolic victory for the mentally ill, benefit design alone cannot address many of the underlying barriers to care. Health benefits mandates do not reach the approximately forty million Americans who are uninsured.[96] In addition, the impact of mandated mental health parity will vary considerably across plans; it will depend on how generous existing benefits are for medical care and whether employers respond to MHETA by expanding mental health benefits to the level of medical and surgical benefits or tightening all benefits across the board.[97] Finally, legislation alone cannot address other barriers to mental health treatment, such as patients' self-awareness, desire for care, understanding of the services available, and fear and stigma

---

[91] SURGEON GENERAL, *supra* note 9, at 411–12.

[92] NATIONAL MENTAL HEALTH ASSOCIATION, *supra* note 16, at 3. An additional $92 billion is spent annually on direct treatment costs. *Id*; *see also* SURGEON GENERAL, *supra* note 9, at 411.

[93] *See* SURGEON GENERAL, *supra* note 9, at 411.

[94] *See, e.g.*, *Mental Health Parity Hearings*, *supra* note 66, at 90–97 (statement of Tara Wooldridge, Manager, Employee Assistance Programs, Delta Air Lines); KIRCHSTEIN, *supra* note 54, at 19–20. This relationship may work in the opposite direction as well, with lower expenditures on mental health services actually lowering worker productivity and increasing absenteeism. *See* Robert A. Rosenheck, et al., *Effect of Declining Mental Health Service Use on Employees of a Large Corporation*, HEALTH AFF., Sept.-Oct. 1999, at 193, 197–99. A study of a large corporation employing over 20,000 workers found that cutbacks in mental health services over a period of several years were associated with a 22% increase in employee sick days. *See id.*

[95] *See* KIRCHSTEIN, *supra* note 54, at 18; *see also* Brian Cuffel, William Goldman & Herbert Schlesinger, *Does Managing Behavioral Health Care Services Increase the Cost of Providing Medical Care?*, J. BEHAV. HEALTH SERVICES & RES. 372, 376–77 (1999). This effect may be due to a tendency of those suffering from mental disorders to perceive their symptoms as physical problems. *See id.* at 373. This tendency, coupled with the limited access to mental health services in a non-parity regime, shifts treatment of the mentally ill into the general medical services sector. *See id.*

[96] ROBERT J. MILLS, U.S. CENSUS BUREAU, CURRENT POPULATION REPORTS: HEALTH INSURANCE COVERAGE 1999 (2000).

[97] *See* Zuvekas, Banthin & Selden, *supra* note 14, at 139–42.

associated with seeking treatment.[98] In the end, mental health parity legislation represents only one incremental step towards ensuring equal access and coverage for all mentally ill individuals.

Opposition to mental health parity, particularly to full parity, has centered on the projected costs of implementation and the resultant effects on both employers and employees. In an efficient and free labor market, it can be expected that employers and insurers will pass the costs of implementing new benefits mandates along to employees.[99] Employers may reduce compensation or other benefits, reduce the level of health insurance benefits provided (either by increasing limits on care or through higher deductibles and co-payments), or drop health insurance coverage entirely.[100] The result is that more employees may lose their health insurance because their employers no longer offer it or because they can no longer afford to pay their share of the costs.[101] Recent studies indicate that one-fourth to one-fifth of uninsured Americans do not have health coverage because of the costs of various health care benefit mandates.[102] Ironically, individuals with poor mental health may be more likely to lose their insurance because they are more likely to be members of vulnerable groups.[103] Dean Rosen, senior vice president of policy and general counsel for the Health Insurance Association of America, outlined these concerns in his testimony before the Senate Committee on Health, Education, Labor and Pensions in May 2000:

---

[98] *See* HAROLD E. VARMUS, NATIONAL INSTITUTES OF HEALTH, PARITY IN FINANCING MENTAL HEALTH SERVICES: MANAGED CARE EFFECTS ON COST, ACCESS, AND QUALITY, AN INTERIM REPORT TO CONGRESS BY THE NATIONAL ADVISORY MENTAL HEALTH COUNCIL 17 (1998).

[99] *See* S. REP. NO. 107-61, at 10–11; *see also* Paul Fronstin, *Issues in Mental Health Care Benefits: The Costs of Mental Health Parity*, ISSUE BRIEF NO. 182 (Employee Benefit Research Institute), Feb. 1997, at 10–12 (1997); JENSEN & MORRISEY, *supra* note 65, at 11.

[100] *See* S. REP. NO. 107-61, at 10–11 (2001); *see also* Fronstin, *supra* note 99, at 10–12; JENSEN & MORRISEY, *supra* note 65, at 11.

[101] *See* S. REP. NO. 107-61, at 10–11; *see also* JENSEN & MORRISEY, *supra* note 65, at 14.

[102] *See* JENSEN & MORRISEY, *supra* note 65, at 14. Federal and state mental health parity mandates are one element in a patchwork of over 1000 health insurance coverage mandates. *See id.* Not surprisingly, studies indicate that the risk of individuals losing their health insurance grows with each additional mandate. *See id.*

[103] *See* Roland Sturm & Kenneth Wells, *Health Insurance May Be Improving—But Not for Individuals with Mental Illness*, HEALTH SERVICES RES., Apr. 2000, at 253, 257–60. A recent survey asking individuals about changes in their health insurance coverage during the past two years found that those with poor mental health indicators were more likely to report a decrease in health benefits or loss of insurance, an inability to regain coverage once it was lost, and poorer access to care. *See id.* These results held even when controlling for socio-demographic factors such as income and level of schooling. *See id.* A similar survey found a 2.4% increase in insurance loss for individuals with signs of mental health disorder in states where parity was implemented. *See* Roland Sturm, *State Parity Legislation and Changes in Health Insurance and Perceived Access to Care Among Individuals with Mental Illness: 1996–1998*, 3 J. MENTAL HEALTH POL'Y & ECON. 209, 211–13 (2000).

> Ultimately, increased costs are not borne by health insurers—they fall on employees and individual consumers. There is a balance that must be achieved between benefit mandates and the cost of insurance—where that balance is struck will have an impact on how many Americans have health coverage, and how many do not.[104]

Along these lines, many criticize benefits mandates for depriving employers and employees of choice in designing benefits and in striking the balance between coverage and cost, processes that they argue are best resolved in the free market.[105]

A counter-argument to the parity proponents' charges of discrimination is that inequalities in mental health insurance coverage are actually caused by two market forces: "adverse selection" and "moral hazard."[106] Adverse selection refers to the tendency of health insurance plans offering the most comprehensive coverage to attract those individuals most in need of care, leading to inflated costs, and causing healthier individuals to switch to cheaper plans offering less coverage.[107] A recent national study found some empirical support for adverse selection related to mental health benefits: employees with family members suffering from mental illness appear to seek out employment positions that offer more comprehensive mental health coverage.[108] The term moral hazard refers to the tendency of insured individuals to over-utilize health care benefits because they do not have to pay for most of the total costs of care under typical health plans, resulting in patients seeking services that are unnecessary, wasteful, and inefficient.[109] While moral hazard is a risk for all forms of health insurance, research suggests that demand for mental health services is even more responsive to changes in cost than demand for general medical services.[110] The result is that health plans must apply

---

[104] *Mental Health Parity* Hearings, *supra* note 66, at 100 (statement of Dean Rosen, Senior Vice President of Policy and General Counsel, Health Insurance Association of America). Rosen cited studies indicating that each 1% increase in health insurance premiums will result in loss of coverage for 200,000 to 300,000 Americans. *See id.*

[105] *See, e.g.*, JENSEN & MORRISEY, *supra* note 65, at 14; 142 CONG. REC. S9922 (1996) (statement of Sen. Gramm). Labor unions have also opposed parity on grounds of freedom of choice, arguing that benefits mandates constrain their options in negotiations with employers. *See* Erica Goode, *Equal Footing*, N.Y. TIMES, Jan. 1, 2001, at A1.

[106] *See* SURGEON GENERAL, *supra* note 9, at 420.

[107] *See id.*

[108] *See* Marc P. Freiman, *Interactions Between Use of and Insurance for Special Ambulatory Mental Health Services*, 1 J. MENTAL HEALTH POL'Y & ECON. 119, 123–26 (1998). This study did not find evidence, however, that employees suffering from mental health disorders themselves were able to seek out positions that provided broader coverage. *See id.* at 123–24. The author suggests that this may be due to a selective bias of employers, with those offering more comprehensive coverage tending not to hire individuals who overtly display signs of mental health problems. *See id.* at 123–26.

[109] *See* SURGEON GENERAL, *supra* note 9, at 420.

[110] *See* Richard G. Frank & Thomas G. McGuire, *Parity for Mental Health and Sub-*

higher cost-sharing mechanisms, such as co-payments and deductibles, to mental health services in order to minimize moral hazard.[111]

The phenomena of moral hazard and adverse selection, however, may cut in opposite directions. Adverse selection suggests that, in an unregulated market, mentally ill individuals will be forced to choose between inadequate but affordable coverage and more comprehensive plans with inflated costs.[112] Only a legal mandate requiring plans to offer some minimal level of coverage can address this market failure.[113] On the other hand, moral hazard provides a justification for applying higher cost-sharing mechanisms to insurance coverage for mental health services in order to minimize consumer abuse and maximize consumer value.[114] According to this model, a broad mandate requiring equal benefits could encourage wasteful and inefficient utilization of mental health services.[115]

Another element of the debate over cost is the question of how a federal law mandating full mental health parity would be affected by other recent developments in the health care industry, including the beginning of a shift away from managed care. As discussed in further detail below, it is the use of managed care and its cost-controlling techniques that are vital to controlling costs under full mental health parity. In his testimony, Dean Rosen argued that current cost projections for mental health parity may underestimate the impact parity would have on increasing costs because of a recent and continuing backlash against managed care in the health insurance market.[116] Public opinion surveys and market data show that consumers are beginning to reject tightly controlled managed care plans, and patient protection bills being considered by the current Congress also would limit certain managed care techniques.[117] If health insurers cannot rely on managed care techniques to control costs under parity, a new federal mandate may be more expensive than anticipated.[118] Senator Judd Gregg (R-N.H.) recently expressed similar concerns about the costs of implementation when viewed in light of other health care reform legislation, the backlash against tightly man-

---

stance Abuse Care Under Managed Care, 1 J. MENTAL HEALTH POL'Y & ECON. 153, 153–56 (1998).

[111] *See id.*

[112] *See Achieving Parity for Mental Health Treatment Hearings, supra* note 51, at 4 (statement of Darrel A. Regier); *see also* SURGEON GENERAL, *supra* note 9, at 420, 423.

[113] *See Achieving Parity for Mental Health Treatment Hearings*, *supra* note 51, at 4 (statement of Darrel A. Regier).

[114] *See* Frank & McGuire, *supra* note 110, at 153–56.

[115] *See id.*

[116] *See Mental Health Parity Hearings, supra* note 66, at 102–03 (statement of Dean Rosen, Senior Vice President of Policy and General Counsel, Health Insurance Association of America).

[117] *See id.* (citing market penetration data indicating that consumers are shifting from more tightly managed health maintenance organizations to more flexible forms of managed care such as preferred provider organizations).

[118] *See id.* at 101–03; *see also* S. REP. NO. 107-61, at 18–19.

aged care, and ongoing double-digit premium increases every year.[119] As both Dean Rosen and Senator Gregg emphasize, it is important to assess the potential impact of new mandates for health plans in tandem with other market and political trends.

In 1996, supporters of federal mental health parity legislation faced significant opposition from business interests, and they ultimately succeeded in passing the MHPA only by scaling back the scope of their original proposal significantly.[120] After the Senate failed to act on an earlier proposal,[121] Senators Domenici and Wellstone introduced provisions requiring full parity for all mental health services as an amendment to Senate Bill 1028, the Health Insurance Reform Act.[122] Although Domenici and Wellstone were able to overcome opposition in the Senate (winning a vote to keep the amendment in the bill by 65 to 33),[123] the provisions requiring mental health parity were dropped during conference committee negotiations.[124]

---

[119] *See* S. REP. NO. 107-61, at 18–19. The small business exemption is, in part, a reaction to the particularly high premium increases faced by small businesses in recent years. *See id.* at 18.

[120] *See generally* Frank, Koyanagi, & McGuire, *supra* note 8; Gitterman, Sturm, Pacùla & Scheffler, *supra* note 45; Levinson & Druss, *supra* note 37.

[121] *See* Equitable Health Care for Serious Mental Illnesses Act of 1995, S. 298, 104th Cong. (1995). This bill would have committed the federal government to a policy of nondiscriminatory and equitable health insurance coverage for serious mental illnesses; it laid the groundwork for future legislation but stopped short of creating any specific new requirements. *See id.*

[122] *See* 142 CONG. REC. S3588–92 (1996). This far-reaching amendment would have required both group and individual health plans to equalize all financial requirements and treatment limitations between coverage of mental health services and medical services. *See id.* Of all prior proposals, this one most closely matches the current proposed Act. The sponsors of the bill, Senators Edward Kennedy (D-Mass.) and Nancy Kassebaum (R-Kan.), fought the amendment because they feared it would become a poison pill that would doom their bill. *See* 142 CONG. REC. S3591 (1996); *see also* Frank, Koyanagi & McGuire, *supra* note 8, at 113; *See* Judith Havemann, *Businesses Oppose Parity For Mental Health Benefits—Provision in Senate Measure Seen as Too Costly*, WASH. POST, Apr. 26, 1996, at A1.

[123] *See* 142 CONG. REC. S3592 (1996). There is some reason to doubt that this vote was an accurate measure of Senate support for full parity. Senator Phil Gramm (R-Tex.) later commented: "We adopted [the mental health parity amendment], but we adopted it when the majority leader, Senator Dole, made it clear that we were never going to see it emerge from conference." 142 CONG. REC. S9922 (1996). Commentators at the time also reacted with surprise to the vote of support for parity. *See, e.g.*, Laura Blumenfeld, *When Politics Becomes Personal—All They Can Agree on is the Pain of Mental Illness*, WASH. POST, June 19, 1996, at C1 ("The senators surprised everyone that night, including themselves, and voted overwhelmingly in favor of the mental health parity amendment.").

[124] *See* Levinson & Druss, *supra* note 37, at 141. The mental health parity provisions were rejected despite a last-minute compromise offered by Senator Domenici that would have dropped coverage for substance abuse and allowed unequal co-payments and deductibles to continue. *See* Judith Havemann & Helen DeWar, *Mental Health Insurance Provision Fights for Life*, WASH. POST, June 5, 1996, at A4. Senators Kennedy and Kassebaum continued to insist that all controversial amendments be removed from the bill in order to improve its chances of final passage. *See* Levinson & Druss, *supra* note 37, at 141.

Following this failure, Domenici and Wellstone met with former Republican Senate Majority Leader Trent Lott (R-Miss.) and former Speaker of the House Newt Gingrich (R-Ga.), and agreed to scale back their proposal significantly.[125] From this point forward, the tide began to turn in favor of enacting a very limited form of parity. Domenici and Wellstone introduced their new proposal as an amendment to the Fiscal Year 1997 appropriations bill for the Departments of Veterans' Affairs and Housing and Urban Development, which had already been approved by the House. The amendment won Senate approval by a vote of 75 to 22.[126] The House then voted 392 to 17 in favor of a non-binding resolution instructing the House conference committee members to support the parity provisions approved by the Senate.[127] The conference committee approved the scaled-back provisions, the appropriations bill was enacted, and the MHPA became law on September 26, 1997.[128]

Opposition to the MHPA centered on the potential costs of the proposal for businesses.[129] Senators Domenici and Wellstone relied on two broad strategies to overcome this opposition. First, throughout the legislative process they demonstrated their willingness to compromise on the scope of the proposal,[130] and ultimately agreed to significantly curtail its provisions.[131] By limiting the proposal to equalizing annual and lifetime dollar limits and adding an exemption for businesses that experienced increased costs, the sponsors were able to answer objections to the potential cost impact of the bill.[132] In addition, Senators Domenici and Wellstone chose to introduce their bill in a manner that forced it onto the legislative agenda while simultaneously minimizing the opportunity for debate, first as an amendment to a broader health reform measure and then as an amendment to an appropriations bill.[133]

---

[125] *See* Levinson & Druss, *supra* note 37, at 141–42.

[126] *See* 142 CONG. REC. S9916–25 (1996).

[127] *See* 142 CONG. REC. H10182–89 (1996).

[128] *See* 142 CONG. REC. H10733–57, H10951–70, H11227, S11218–25, S11290 (1996).

[129] Among the industry groups lobbying against the parity provisions were the Chamber of Commerce, the National Association of Manufacturers, the American Association of Private Pension and Welfare Plans, and the ERISA Industry Committee. *See* Havemann, *supra* note 122, at A1. An umbrella group representing 130 of the largest employers in the United States publicly withdrew its support for the Health Insurance Reform Act because of the mental health parity requirements. *See id.* Cost was the primary reason cited by these groups for opposing the proposal. *See id.*; *see also* Havemann & Dewar, *supra* note 124, at A4.

[130] *See, e.g.*, Levinson & Druss, *supra* note 37, at 140 (quoting Senator Domenici as saying, "I don't want anyone to go out of business because of a mandate we put on them. If they think the first bite of this apple is too big, let's talk.").

[131] *See, e.g., id.*

[132] *See id.* (noting that revisions to the bill made it "acceptable to its adversaries").

[133] *See id.* at 140–41. For example, during the debate over the Health Insurance Reform Act in the Senate, opponents of the mental health parity measure were more focused on other, more significant amendments being introduced at the same time. *See id.*; *see also* Frank, Koyanagai & McGuire, *supra* note 6, at 113.

Similar strategies have unfolded in the Congressional debate in 2001 over the MHETA. As in 1996, the bill was introduced as an amendment to a broader appropriations bill.[134] Another similarity is that a number of business groups stand opposed to the measure, based on its potential costs, and they are lobbying Congress to prevent the bill's passage.[135] Although the concerns about cost remain, both supporters and opponents of the current bill seem to agree that "the proposal stands a greater chance of enactment now than at any other time in the last decade."[136] The reason may be that supporters have been able to effectively counter opposition to the bill by marshalling recent data indicating that federally mandated mental health parity would have only a minimal impact on health insurance costs.

Understanding of the cost implications of mental health parity has progressed significantly since 1996, shaped by several key developments: the increased role of managed care, a better understanding of how managed care practices can control costs under full parity, and the new availability of data from actual experience with parity that demonstrate its feasibility under managed care. The most important change during the past five years has been the penetration of managed care into the health insurance market, particularly in the area of mental health services.[137] Approximately two-thirds of Americans covered by employer-based insurance plans now receive their care through either a health maintenance organization or a preferred provider organization.[138] Moreover, a majority of employers and public insurance systems now rely on managed behavioral health organizations to manage their mental health benefits sepa-

---

[134] It appeared that the MHETA might be neglected in the aftermath of the September 11 terrorist attacks and the resultant shift in Congress's legislative agenda. *See, e.g., Advocates Are Respectfully Trying To Keep Federal Parity Legislation From Lapsing After Terrorist Assault in New York and Washington, D.C.*, MENTAL HEALTH WKLY., Sept. 24, 2001, at 1.

[135] *See* Robert Pear, *Furious Lobbying Is Set Off By Bill On Mental Health*, N.Y. TIMES, Nov. 6, 2001, at A1; *see also* Am. Ass'n. of Health Plans, *Washington Bulletin October 1, 2001: Crowded Legislative Agenda Includes Room for Health Issues* (2001), *at* http://www.aahp.org (last visited Nov. 15, 2001); Letter from Mark J. Ugoretz, President, ERISA Industry Committee, and Anthony J. Knettel, Vice-President Health Affairs, ERISA Industry Committee, to all Labor—HHS Appropriations Conferees (Nov. 12, 2001), *available at* http://www.eric.org (last visited Nov. 15, 2001).

[136] *See* Pear, *supra* note 135.

[137] *See generally* SURGEON GENERAL, *supra* note 9, at 422–23.

[138] *See* EMPLOYEE BENEFIT RESEARCH INSTITUTE, EMPLOYMENT-BASED HEALTH CARE BENEFITS AND SELF-FUNDED EMPLOYMENT-BASED PLANS: AN OVERVIEW 4 (2000); *see also* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 15 (reporting from employers surveyed in 1999 that 89% of their most popular health plans contained managed care features, and that fourteen contained more managed care features in 1999 compared to 1996). By comparison, one national survey found that fee-for-service plans were still the most prevalent type of plan until as recently as 1992, covering approximately 62% of employees at that time. VARMUS, *supra* note 98, at 8.

rately from their medical benefits; these managed organizations provide coverage for approximately 170 million Americans.[139]

Managed care presents a potential solution to many of the problems raised by opponents of mental health parity, but it also may require a fundamental re-thinking of what parity means. Managed care controls costs through a variety of supply-side techniques, such as pre-certification and ongoing utilization review, the use of networks of providers with pre-negotiated (lower) fees, and a shift to less-intensive (and less-expensive) forms of treatment.[140] The introduction of managed care thus replaces demand-side techniques, such as treatment limitations and financial requirements—which are eliminated or sharply restricted under parity—and offers an alternative mechanism for controlling costs.[141] Managed care techniques such as pre-certification and utilization review respond directly to the problem of moral hazard, by ensuring that only necessary and appropriate care is provided.[142] Allowing managed behavioral health organizations to carve out mental health benefits and pool individuals from multiple health plans also may address the problem of adverse selection.[143] The challenge raised by managed care of mental health services, however, is that supply-side restrictions make benefit design less important.[144] While parity under a traditional fee-for-service model only required that mental health services would be subjected to the same demand-side restrictions as other medical services, parity in the

---

[139] *See Achieving Parity for Mental Health Treatment Hearings*, *supra* note 51, at 1 (statement of Henry Harbin, M.D., Chairman and CEO, Magellan Health Services); *see also* SURGEON GENERAL, *supra* note 9, at 144. Under a "carve-out" plan, an employer or insurer subcontracts with a third party, a managed care organization specializing in mental health benefits, to manage these benefits separately from all other health care benefits. *See* SURGEON GENERAL, *supra* note 9, at 421. According to one estimate, the number of individuals covered by behavioral health care carve-out plans doubled between 1992 and 1998. *See* Richard M. Scheffler, *Managed Behavioral Health Care and Supply-Side Economics*, 2 J. MENTAL HEALTH POL'Y & ECON. 21, 22 (1999). For a broad discussion of the economic advantages of managed behavioral health care carve-outs over traditional fee-for-service and managed care plans, see generally Ingo Vogelsang, *Economic Aspects of Mental Health Carve-Outs*, 2 J. MENTAL HEALTH POL'Y & ECON. 29 (1999).

[140] *See* SURGEON GENERAL, *supra* note 9, at 423; *see also* Scheffler, *supra* note 139, at 22–23; Vogelsang, *supra* note 139, at 32–37. Under pre-certification and utilization review procedures, managed care providers require prior and ongoing approval of treatment in order to ensure that services are limited to necessary and appropriate care. *Cf.* SURGEON GENERAL, *supra* note 9, at 422–23; Vogelsang, *supra* note 139, at 33.

[141] *See* SURGEON GENERAL, *supra* note 9, at 423; *see also* Scheffler, *supra* note 139, at 22–23; Vogelsang, *supra* note 139, at 32–37.

[142] *See* SURGEON GENERAL, *supra* note 9, at 422–23; Vogelsang, *supra* note 139, at 33.

[143] *See* Vogelsang, *supra* note 139, at 36. Recall that adverse selection occurs when the market offers multiple plans with varying degrees of coverage, and individuals most in need of care gravitate towards plans offering the most generous benefits, causing increased costs in those plans. *See supra* text accompanying notes 110–112. By providing uniform coverage to all beneficiaries and combining individuals in a large pool of shared risk, managed behavioral health plans prevent individuals from choosing health plans based on their coverage of mental health benefits. *See* Vogelsang, *supra* note 139, at 36.

[144] *See* Frank & McGuire, *supra* note 110, at 156–58.

new managed care environment will also require that supply-side restrictions are applied equally as well.[145]

The ability of managed care (and in particular managed behavioral health organizations) to provide alternative cost controls in conjunction with parity for mental health services has been confirmed by a number of recent case studies. States that have carved out mental health services in their state employees' health benefits plans and expanded benefits at the same time have seen overall mental health expenditures fall by up to 50 or 60%.[146] Large employers carving out and expanding mental health benefits have similarly experienced comparable cost declines.[147] For example, one employer with over 179,000 insured employees saw its average costs per member per month decline 43% over a six-year period after implementing a mental health carve-out with expanded benefits.[148] This same employer had been experiencing annual increases in mental health costs of twenty to 30% prior to its change in benefits.[149] The cost savings documented in these case studies were achieved primarily through lower pre-negotiated costs of care[150] and dramatic declines in inpatient spend-

---

[145] *See id.*

[146] *See, e.g.*, VARMUS, *supra* note 98, at 10–12 (reporting cost decreases of over 50% over a five-year period for Texas state employees plan, and 32% in five years for North Carolina state employees plan); Ching-to A. Ma & Thomas G. McGuire, *Cost and Incentives in a Behavioral Carve-Out*, HEALTH AFF., Mar.-Apr. 1998, at 53, 63–65 (finding overall cost decline of 57% over four-year period for Massachusetts state employees plan, with at least 30 to 40% attributed solely to the carve-out); Roland Sturm, William Goldman & Joyce McCulloch, *Mental Health and Substance Abuse Parity: A Case Study of Ohio's State Employee Program*, 1 J. MENTAL HEALTH POL'Y & ECON. 129, 132 (1998) (showing significant and continuing decline in costs per member per month over seven year period in Ohio state employees plan). *Cf.* James J. Callahan, Donald S. Shepard, Richard H. Beinecke, Mary Jo Larson & Doreen Cavanaugh, *Mental Health/Substance Abuse Treatment in Managed Care: The Massachusetts Medicaid Experience*, HEALTH AFF., Fall 1995, at 173, 176, 178 (reporting decline of 27% in mental health expenditures over four years, amounting to 19% decrease in costs per enrollee in Massachusetts Medicaid program covering approximately 375,000 lives).

[147] *See, e.g.*, KIRCHSTEIN, *supra* note 54, at 12 (finding decrease in per member costs of 50% after three years for employer covering over 150,000 lives); Cuffel, et al., *supra* note 95, at 376–78 (reporting a 68% decline in average costs over a four-year period for Alcan, covering approximately 13,000 lives). A comprehensive study of thirty employers over a period of seven years found that the introduction of managed behavioral care carve-outs contributed to a continuous average decline of $0.28 per member per month in overall costs. *See* William Goldman, Joyce McCulloch, Brian Cuffel & Danah Kozma, *More Evidence for the Insurability of Managed Behavioral Health Care*, HEALTH AFF., Sept.-Oct. 1999, at 172, 176.

[148] William Goldman, Joyce McCulloch & Roland Sturm, *Costs and Use of Mental Health Services Before and After Managed Care*, HEALTH AFF., Mar.-Apr. 1998, at 40, 41–42, 45–46.

[149] *See id.*

[150] *See id.* at 46–47 (reporting decline in costs charged to the health plan of 40% per inpatient day, where inpatient day cost reductions are one of the main factors behind overall cost reductions); *see also* Callahan, et al., *supra* note 146, at 177–78; Ma & McGuire, *supra* note 146, at 65.

ing,[151] with a shift in care from inpatient to intermediate and outpatient settings.[152]

The congressional debate over the enactment of the MHPA in 1996 and the current debate over the MHETA illustrates the importance of cost estimates to the political environment surrounding parity legislation. As parity was debated during the 104th Congress, a number of studies were quickly produced to estimate the costs of implementation. A 1996 study by the Congressional Budget Office ("CBO") found that full parity would increase health insurance premiums by 4% overall and cause 400,000 workers (plus 400,000 of their dependents) to lose coverage.[153] Five private studies conducted at the same time estimated that premium increases could range from 3.2% to 8.7% for full parity for all mental illnesses, and from 2.5% to 11.4% for full parity limited to serious mental illnesses.[154]

By contrast, all recent estimates have predicted that full parity in mental health benefits, mandated at the federal level, would increase premiums by only 1 or 2%. In June 2000, the National Advisory Mental Health Council reported that full parity would add only 1.4% to premiums.[155] The CBO now estimates that MHETA would result in a modest

---

[151] *See* VARMUS, *supra* note 98, at 10 (reporting drop of over 50% in inpatient costs in Texas state employees plan); *see also* Callahan, et al., *supra* note 146, at 176–77 (finding that inpatient care accounted for 78% of total cost savings, due to a 7.2% decline in inpatient admissions and a 12.3% decline in length of stay); Goldman, McCulloch & Sturm, *supra* note 148, at 46–47 (reporting that inpatient care fell from 51% to 20% of total mental health costs, due to declines in inpatient admission rate and length of stay).

[152] *See* KIRCHSTEIN, *supra* note 54, at 12; *see also* Goldman, McCulloch & Sturm, *supra* note 148, at 46–47; Sturm, Goldman & McCulloch, *supra* note 146, at 132. Some of these studies have shown that, while outpatient service use and costs have increased under managed care carve-outs, increases have been more than offset by declines in inpatient spending. *See, e.g.*, KIRCHSTEIN, *supra* note 54, at 12 (reporting that increases in outpatient service use and costs were offset by declines in inpatient spending by a factor of four or five to one in carve-out for large employer).

[153] *See* Louis de la Parte Mental Health Institute, *CBO's Estimates of the Impact on Employers of the Mental Health Parity Amendment in H.R. 3103*, at 2, 4 (May 1996), *at* http://www.fmhi.usf.edu/parity/parityhome.html. The CBO predicted that full parity would impose costs of $11.6 billion annually on the private sector, and that employees would lose almost $6 billion in wages and other benefits as employers shifted these increased costs onto their staff. *See id.* at 3. Compare this to the CBO's final estimate that the enacted version of the MHPA would raise premiums by only 0.4% (0.16% after accounting for employee and employer responses). GENERAL ACCOUNTING OFFICE, *supra* note 25, at 17; *see also* MERRILE SING, STEVEN HILL, SUZANNE SMOLKIN & NANCY HESIER, U.S. DEP'T OF HEALTH & HUMAN SERVS., THE COSTS AND EFFECTS OF PARITY FOR MENTAL HEALTH AND SUBSTANCE ABUSE INSURANCE BENEFITS 20 (1998), *available at* http://www.mentalhealth.org/cmhs/ManagedCare/Parity/prtyfnix.asp.

[154] *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 17–18; *see also* SING, et al., *supra* note 153, at 20. The studies were performed by Coopers and Lybrand (3.2% for all mental illnesses), Milliman and Robertson (3.9% for all mental illnesses and 2.5% for serious mental illnesses only), Price Waterhouse(8.7% for all mental illnesses), and Watson Wyatt Worldwide (11.4% for serious mental illnesses only). *See* GENERAL ACCOUNTING OFFICE, *supra* note 25, at 17–18; *see also* SING, et al., *supra* note 153, at 20.

[155] KIRCHSTEIN, *supra* note 54, at 10–11, 32–35.

premium increase of 0.9% across all plans, only 0.4% after accounting for responses by employers and insurers.[156] The MHETA provisions are almost the same as those used in the 1996 CBO analysis of the MHPA as originally passed, and yet during those five years the estimated impact on insurance premiums has dropped by a factor of four.[157] Another recent analysis of the MHETA provisions by PricewaterhouseCoopers estimates that the bill would result in premium increases of 1%, or $1.32 per member per month.[158] The Office of Personnel Management reports that full mental health parity has increased premiums in FHEBP by 1.3% overall.[159] Actual industry experience under managed behavioral health care also supports these smaller figures. Henry Harbin of Magellan Health Systems testified in the Senate earlier this year that his company's clients generally experience premium increases of under 1%.[160] During the debate on the MHETA amendment, supporters cited to this recent data to argue that the cost impact of the bill will be minimal.[161]

Even if Congress enacts a broad federal mandate requiring full parity for mental health services, questions remain about the impact of man-

---

[156] *See* S. REP. No. 107-61, at 10–11. Predicted responses to the new law include fewer employers offering health insurance coverage or coverage of mental health benefits, employers restricting the scope of benefits offered or increasing premiums charged to employees, and fewer employees enrolling in health insurance plans because of higher costs. *See id.* The CBO predicts that these responses would offset 60% of the potential costs of the law. *See id.* The total direct costs to the private sector would be approximately $2.8 billion in 2002, rising to $5.0 billion annually by 2006. *Id.* at 12–13. It is expected that employers would pass on the majority of the final costs of implementation to their employees through lower compensation and benefits, resulting in a loss of approximately $1 billion of taxable income in 2002, growing to $2.3 billion in taxable income annually by 2011. *See id.* at 11.

[157] *See* Louis de la Parte Mental Health Institute, *supra* note 153, at 2; *see also* S. REP. No. 107-61, at 10–11. The divergence in the results of these various studies can be attributed to widely varying assumptions about penetration of managed care in the market, expected consumer responses and demand, and costs of care. *See* KIRCHSTEIN, *supra* note 54, at 10–12, 32–35. All of the government estimates have relied on different versions of a single model developed by the Hay Group. *See id.* at 10–12, 32–35; see also S. REP. No. 107-61, at 10. The most recent version of the model contains data from the FEHBP, several large managed behavioral care companies, and one state employees health plan. *See* KIRCHSTEIN, *supra* note 54, at 11. The older versions of this model relied on historical data and did not reflect current practices under managed care. *See id.* at 32–35; *see also* VARMUS, *supra* note 98, at 6, 13–14; Roland Sturm, *How Expensive is Unlimited Mental Health Care Coverage Under Managed Care?*, 278 J. AM. MED. ASS'N 1533, 1534–36 (1997).

[158] *See* Press Release, American Psychological Association Practice Organization, Mental Health Parity Coverage to Cost $1.32 per Month (Aug. 1, 2001), *at* www.apa.org/practice/paritycoverage.html.

[159] *See Achieving Parity for Mental Health Treatment Hearings, supra* note 51, at 8 (statement of William E. Flynn, III).

[160] *See id.* at 4 (statement of Henry Harbin). Magellan Health Systems is one of the nation's leading providers of managed behavioral health care, serving 70 million individuals. *See id.*

[161] *See* 147 CONG. REC. S11,666 (daily ed. Oct. 30, 2001) (statement of Sen. Domenici); *see also* 147 CONG. REC. S11,169 (daily ed. Oct. 30, 2001) (statement of Sen. Kennedy).

aged care on access to and quality of care. Case studies of individual experiences with managed care and expanded benefits indicate that cost savings can be achieved by service providers without threatening access to care. Even in some plans achieving significant overall cost savings, the percentage of enrollees receiving mental health services each year (referred to as "treated prevalence") has increased.[162] Other evidence casts doubt on these findings, however, suggesting that mental health parity laws have no noticeable impact on treated prevalence rates. A comparative study of states with and without parity legislation found that the presence of parity laws had no effect on the percentage of individuals receiving mental health services or the average number of specialty visits for users.[163] Even when the study focused on those states providing the most comprehensive parity mandates, parity legislation had no significant effect on the general population, but may have had a small positive effect on the number of specialty visits by individuals in poor mental health.[164]

In addition to these unanswered questions about access, there has been very little focus to date on the long-term impact of managed care on the quality of mental health services, and the few studies addressing these issues indicate cause for concern. For example, a study of a managed care mental health carve-out for Medicaid enrollees in Massachusetts found that the thirty day readmission rates for children receiving mental health services under this plan increased 10.1% following the carve-out.[165] Data from another managed care plan show that mental health services were targeted more intensely for reduced inpatient stays than general medical services, and each reduced inpatient day was associated with a 3.1% rise in patients' sixty-day readmission rates.[166] Outcomes under managed care may be even poorer for those with serious mental illnesses.[167] These results raise questions about the long-term vi-

---

[162] *See* KIRCHSTEIN, *supra* note 54, at 12 (reporting increase in treated prevalence from 4.9% to 7.3%); Callahan, et al., *supra* note 146, at 179–80 (reporting slight increase in overall treated prevalence, and increase of 10.6% among enrollees receiving outpatient treatment); Cuffel, et al., *supra* note 95, at 377 (reporting increase in treated prevalence from 5.2% to 6.1%); Goldman, McCulloch & Sturm, *supra* note 148, at 46. *But see* Goldman, McCulloch, Cuffel & Kozma, *supra* note 147, at 176 (reporting no significant change in treatment prevalence and decreases in both inpatient and outpatient sessions per user); Ma & McGuire, *supra* note 146, at 66 (showing decline in both percentage of enrollees treated and number of sessions per user for outpatient care).

[163] *See* Rosalie Liccardo Pacula & Roland Sturm, *Mental Health Parity Legislation: Much Ado About Nothing?*, HEALTH SERVICES RES., Apr. 2000, at 263, 270.

[164] *See id.* at 270–71.

[165] *See* Callahan, et al., *supra* note 146, at 180–81. However it should be noted that the thirty-day readmission rate dropped slightly for the overall Medicaid population. *See id.*

[166] *See* David Mechanic & Donna D. McAlpine, *Mission Unfulfilled: Potholes on the Road to Mental Health Parity*, HEALTH AFF., Sept.-Oct. 1999, at 7, 11–13. The data also indicated that denials of care were applied uniformly across mental health disorders, with no noticeable variation based on the seriousness of the diagnosis. *See id.* The authors refer to this as the "democratization" of services under managed care, meaning individuals receive similar levels of treatment despite differences in their diagnoses and needs. *See id.*

[167] *See id.*

ability of managed care for achieving cost savings and quality health care for mental health services.

If the goal of the mental health parity movement is "assuring access to efficacious treatments for mental illness,"[168] then the passage of the MHETA may be viewed as just one step towards this ultimate end. Even with a federal mandate requiring equal coverage in benefit design, however, challenges will remain to ensure full and equal access. In the future, practitioners will need to develop reliable measures for assessing the quality of mental health services, to set benchmark standards for access and quality, and then to implement strategies for using this information to actually improve the delivery of care.[169] As one recent commentary observed: "The challenge for the coming decade is to develop clear standards based on the best evidence and clinical judgment so that parity has substance in implementation as well as in concept."[170]

Federal legislation constraining managed care techniques, for example through patient protection legislation, is another potential response.[171] One of the ironies in the parity debate is that the rise of managed care has made the widespread enactment of parity legislation possible while simultaneously decreasing its potential impact and relevance on actual access to care.[172] Whatever happens at the federal level, the policy issue of the future will be how to ensure that the mentally ill receive truly equal access to high-quality, appropriate services in a managed care environment.

—Beth Mellen Harrison

---

[168] S. REP. No. 107-61, at 15.

[169] *See* KIRCHSTEIN, *supra* note 54, at 16–17; *see also* VARMUS, *supra* note 98, at 20–21, 24–35; Mechanic & McAlpine, *supra* note 166, at 15–18.

[170] Mechanic & McAlpine, *supra* note 166, at 17.

[171] The American Psychological Association, joined by a number of professional mental health organizations, has proposed a mental health patient's bill of rights. *See* AMERICAN PSYCHOLOGICAL ASSOCIATION, MENTAL HEALTH PATIENT'S BILL OF RIGHTS, *at* http://www.apa.org /pubinfo/rights/ (last visited Nov. 15, 2001).

[172] *See generally* OTTEN, *supra* note 18; Burnam & Escarce, *supra* note 24; Frank & McGuire, *supra* note 110; Gitterman, Sturm & Scheffler, *supra* note 26.