

# Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)

Cass R. Sunstein,<sup>†</sup> Daniel Kahneman,<sup>††</sup> and David Schkade<sup>†††</sup>

## CONTENTS

I. INTRODUCTION: ARBITRARINESS, UNPREDICTABILITY, AND VALUATION PROBLEMS . . . . .	2074
A. <i>Two Goals</i> . . . . .	2074
B. <i>Shared Outrage and Erratic Awards</i> . . . . .	2075
II. DETERRENCE, RETRIBUTION, AND THE CONSTITUTION . . . . .	2081
A. <i>The Goals of Punitive Damages</i> . . . . .	2081
1. <i>Deterrence</i> . . . . .	2082
a. <i>Conventional Arguments</i> . . . . .	2082
b. <i>Puzzles and Problems with the Deterrence Justification</i> . . . . .	2084
2. <i>Retribution</i> . . . . .	2085
B. <i>The Constitution and Punitive Damages</i> . . . . .	2086
1. <i>Substantive Versus Procedural Interpretations of Unconstitutional Punishment</i> . . . . .	2086
2. <i>Leading Cases</i> . . . . .	2088
3. <i>The Need for an Understanding of Jury Variability To Inform Constraints on Punitive Damages</i> . . . . .	2092
III. WHY PUNITIVE DAMAGE AWARDS ARE ERRATIC . . . . .	2094
A. <i>Shared Outrage</i> . . . . .	2095
B. <i>Hypotheses</i> . . . . .	2097

---

<sup>†</sup> Karl N. Llewellyn Distinguished Service Professor, Law School and Department of Political Science, University of Chicago.

<sup>††</sup> Eugene Higgins Professor of Psychology and Professor of Public Affairs, Princeton University.

<sup>†††</sup> Professor of Management and William M. Spriegel Fellow, Graduate School of Business, University of Texas, Austin. The authors are grateful to the Exxon Corporation for financial support of the empirical study contained in this Article; Exxon bears no responsibility for our analysis or our conclusions. We are also grateful to Richard Craswell, Richard Epstein, Dan Kahan, William Landes, Tracey Meares, Martha Nussbaum, Eric Posner, and Richard Posner for valuable comments. Thanks also to participants in the rationality workshop at the University of California, Berkeley, the Behavioral Law and Economics Conference at the National Bureau for Economic Research, and the law-and-economics and legal theory workshops at the University of Chicago. J.B. Heaton provided research assistance and excellent comments.

C.	<i>Results</i> . . . . .	2097
1.	<i>Shared Outrage</i> . . . . .	2097
2.	<i>Unpredictable Dollar Awards</i> . . . . .	2100
3.	<i>The Effects of Context, Harm, and Firm Size</i> . . . . .	2104
D.	<i>Discussion</i> . . . . .	2105
1.	<i>The Underlying Problem: "Scaling Without a Modulus"</i> . . . . .	2106
2.	<i>A Simple Way To Improve Predictability</i> . . . . .	2107
IV.	POLICY REFORMS: COMMUNITY JUDGMENTS WITH MEANING AND WITHOUT NOISE? . . . . .	2109
A.	<i>Preliminary Observations on Limitations and Reform</i> . . . . .	2109
1.	<i>Experiments and the Real World</i> . . . . .	2109
2.	<i>Anchors and Their Effects</i> . . . . .	2109
3.	<i>The Rule of Law and How To Obtain (and How Not To Obtain) Its Virtues</i> . . . . .	2110
4.	<i>Deterrence and Retribution</i> . . . . .	2111
B.	<i>Punitive Damages Reform</i> . . . . .	2112
1.	<i>Predictable Populism: Community Judgments About Dollars</i> . . . . .	2114
a.	<i>Administrative Issues</i> . . . . .	2115
b.	<i>Two Phases</i> . . . . .	2115
c.	<i>Using Scenarios and Cases: A Note on Existing Practice</i> . . . . .	2118
2.	<i>Technocratic Populism: Using Punitive Intent but Not Dollars</i> . . . . .	2120
a.	<i>Phase One: The Jury's Role</i> . . . . .	2120
b.	<i>Calibration and Phase Two: Translating Jury Judgments into Dollars</i> . . . . .	2121
3.	<i>Civil Sentencing and Bureaucratic Rationality: The Partial or Complete Elimination of Juries</i> . . . . .	2121
4.	<i>Evaluating the Alternatives: The Proposed Reforms, Caps, and Multipliers</i> . . . . .	2126
V.	IMPLICATIONS, EXTENSIONS, AND SPECULATIONS . . . . .	2131
A.	<i>Difficult Damage Determinations</i> . . . . .	2131
1.	<i>Pain and Suffering</i> . . . . .	2131
2.	<i>Libel</i> . . . . .	2133
3.	<i>Intentional Infliction of Emotional Distress and Sexual Harassment</i> . . . . .	2134
4.	<i>Compensatory Versus Punitive Damages: General Considerations</i> . . . . .	2135
B.	<i>Other Valuation Problems</i> . . . . .	2138
1.	<i>Regulatory Expenditures</i> . . . . .	2138

2. <i>Prison Terms</i> .....	2139
3. <i>Contingent Valuation</i> .....	2140
VI. CONCLUSION .....	2142
APPENDIX A: EXPERIMENTAL DESIGN .....	2146
A. <i>Sample and Procedure</i> .....	2146
B. <i>Design and Stimuli</i> .....	2146
APPENDIX B: EXCERPTS FROM INSTRUCTIONS .....	2147
APPENDIX C: PERSONAL INJURY SCENARIOS .....	2147
APPENDIX D: TABLES .....	2152

I. INTRODUCTION: ARBITRARINESS, UNPREDICTABILITY,  
AND VALUATION PROBLEMS

A. *Two Goals*

This Article has two goals. The first is to investigate the sources of arbitrariness and unpredictability in punitive damage awards. On the basis of responses from 899 jury-eligible citizens, we estimate the results of deliberations consisting of all white juries, all African-American juries, all female juries, all male juries, all wealthy juries, all poor juries, and juries of widely diverse degrees of age and education.

Our principal conclusions, stated briefly,<sup>1</sup> are that at least in the personal injury cases studied here, people's moral judgments are remarkably widely shared, but that people have a great deal of difficulty in mapping such judgments onto an unbounded scale of dollars. Shared moral judgments but erratic, unpredictable, and arbitrary awards, possibly even meaningless awards, are a potential product of this difficulty. Since participants in law are frequently asked to map their judgments onto an unbounded dollar scale, this problem relates not only to punitive damage reform but also to a number of other positive and normative questions now faced by the legal system, including the law governing awards for pain and suffering, libel, sexual harassment and other civil rights violations, intentional infliction of emotional distress, administrative penalties, and contingent valuation. In addition to identifying serious problems in current practices and suggesting possible reforms, our study points in the direction of a large and potentially fruitful research agenda.

Our second goal involves what might be called the behavioral analysis of law. In the past three decades, a great deal of progress has come from the application of a certain understanding of economics to legal problems. Within economics and psychology, but outside law, that understanding has been under sustained attack. Within social science generally, the attack has produced insights that supplement, and sometimes undermine, those versions of economics that have undergirded economic analysis of law. These insights very much bear on law. Our study is thus an effort to contribute to the behavioral analysis of law, a field at its inception but showing signs of considerable growth.<sup>2</sup> Our findings also help show what might be missing, impractical, or

---

1. A technical version of the survey and study, accompanied with more detailed statistical analysis, is reported in Daniel Kahneman, David Schkade, & Cass R. Sunstein, *Shared Outrage and Erratic Awards: The Psychology of Punitive Damages*, 16 J. RISK & UNCERTAINTY (forthcoming Apr. 1998). This Article provides a nontechnical summary of our findings and develops, as the technical paper does not, the implications of our findings for law and legal reform.

2. See, e.g., Robert C. Ellickson, *Bringing Culture and Human Frailty to Rational Actors: A Critique of Classical Law and Economics*, 65 CHI.-KENT L. REV. 23 (1989); Christine Jolls, Cass R. Sunstein & Richard Thaler, *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. (forthcoming July 1998);

wrong in standard economic approaches to punitive damage awards and does so in a way that bears on a range of additional issues as well.

### B. *Shared Outrage and Erratic Awards*

The awarding of punitive damages has become one of the most controversial and important uses of tort law, extending well beyond the common law to such statutory areas as environmental protection<sup>3</sup> and employment discrimination.<sup>4</sup> Punitive damages are allowed in forty-seven of the fifty states,<sup>5</sup> and over sixty federal statutes now permit the award of punitive damages,<sup>6</sup> making judicial review of punitive awards a significant part of federal law.

The purposes of such awards are not obscure. Sometimes compensatory awards provide insufficient deterrence of private behavior because some injured parties do not detect and seek compensation for their injuries. Punitive damages can, in theory, take account of the infrequency of private suits by penalizing defendants enough ex post that they will undertake optimal precautions ex ante.<sup>7</sup> Moreover, punitive damages may have a retributive or expressive function, designed to embody social outrage at the actions of serious wrongdoers.<sup>8</sup> They may reflect the "sense of the community" about the egregious character of defendants' actions.<sup>9</sup>

Whatever their ultimate purposes, the most widespread concern about punitive damages has been that they are unpredictable, even "out of control."<sup>10</sup> Consider, as possible examples of this phenomenon, a punitive

Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. (forthcoming Apr. 1998); Cass R. Sunstein, *Behavioral Analysis of Law*, 64 U. CHI. L. REV. 1175, 1193 (1997).

3. See, e.g., 42 U.S.C. § 9607(c)(3) (1994).

4. See, e.g., *id.* § 1981a(b)(1).

5. See RICHARD L. BLATT ET AL., *PUNITIVE DAMAGES: A STATE-BY-STATE GUIDE TO LAW AND PRACTICE* § 8.2, at 114 tbl.8-1 (1991).

6. See, e.g., 7 U.S.C. § 18(a)(1)(B) (1994); *id.* § 21(b)(10)(C)(ii); *id.* § 25(a)(3)(B); 10 U.S.C. § 2207(a)(2) (1994); *id.* § 2409(c)(2); 15 U.S.C. § 1681n(2) (1994); *id.* § 1681u(i); *id.* § 2622(b)(2)(B)(iv); 18 U.S.C. § 248(c)(1)(B) (1994); *id.* § 2724(b)(2); 42 U.S.C. § 300j-9(i)(2)(B)(ii)(IV); *id.* § 1981a(b)(1); *id.* § 3613(c)(1); *id.* § 7622(d); *id.* § 9607(c)(3); *id.* § 13,981(c).

7. See WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 160-65, 184-85, 223-24 (1987); A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 873-74 (1998).

8. See Marc Galanter & David Luban, *Poetic Justice: Punitive Damages and Legal Pluralism*, 42 AM. U. L. REV. 1393, 1432-40 (1993); Jean Hampton, *The Retributive Idea*, in FORGIVENESS AND MERCY 111, 122-43 (Jeffrie G. Murphy & Jean Hampton eds., 1988); David G. Owen, *The Moral Foundations of Punitive Damages*, 40 ALA. L. REV. 705, 705 (1989).

9. Understood in these terms, however, punitive awards raise several puzzles. Such awards interact in complex ways with public law—the elaborate network of regulatory requirements and criminal prohibitions that overlap with the tort law. Like punitive damages, the requirements and prohibitions of the public law also have deterrent, retributive, and expressive functions. It is not clear how they can best be brought together with tort law to produce a coherent whole.

10. John Calvin Jeffries, Jr., *A Comment on the Constitutionality of Punitive Damages*, 72 VA. L. REV. 139, 139 (1986).

award of \$4 million for nondisclosure of the fact that the plaintiff's new BMW had been repainted;<sup>11</sup> a \$6 million punitive award for tortious interference with contractual relations;<sup>12</sup> a punitive award of \$80 million, reduced to \$7.8 million, for racial discrimination even though the employee had found comparable employment;<sup>13</sup> a \$30 million punitive award, subsequently reduced to \$6 million, for anticompetitive conduct;<sup>14</sup> and a \$2.7 million punitive award, later reduced to \$480,000, to a woman who spilled hot coffee on herself.<sup>15</sup>

These are mere anecdotes, but there is more systematic evidence of unpredictability as well.<sup>16</sup> One study of forty-seven counties in eleven states over a several-year period showed a high degree of variability: Punitive damages were awarded in about 25% of the successful verdict cases in some counties and not awarded at all in others.<sup>17</sup> Median verdicts ranged from less than \$10,000 in some areas to as much as \$204,000 in San Diego.<sup>18</sup> Leaders of the House of Representatives have said, in a way that captures the conventional wisdom, that the arbitrary character of punitive damage awards produces an affront to the rule of law by "distributing awards in a random and capricious manner,"<sup>19</sup> and that they should be subject to a cap of three times the actual harm.<sup>20</sup> Congress continues to debate bills that would limit the availability and amount of punitive damages.<sup>21</sup>

11. See *BMW of N. Am., Inc. v. Gore*, 646 So. 2d 619, 622 (Ala. 1994), *rev'd*, 116 S. Ct. 1589 (1996). The Supreme Court invalidated this award on due process grounds.

12. See *Browning-Ferris Indus., Inc. v. Kelco Disposal, Inc.*, 492 U.S. 257, 259 (1989).

13. See *Lane v. Hughes Aircraft Co.*, 65 Cal. Rptr. 2d 889, 912 (Ct. App.), *review granted and opinion superseded*, 946 P.2d 816 (Cal. 1997).

14. See *Continental Trend Resources, Inc. v. OXY USA Inc.*, 101 F.3d 634, 643 (10th Cir. 1996), *cert. denied*, 117 S. Ct. 1846 (1997).

15. See *Liebeck v. McDonald's Restaurants, P.T.S., Inc.*, No. CV-93-02419, 1995 WL 360309, at \*1 (D.N.M. Aug. 18, 1994); see also *Greene v. Boddie-Noell Enters., Inc.*, 966 F. Supp. 416, 418 n.1 (W.D. Va. 1997) (providing further details about the *Liebeck* case).

16. See Peter Huber, *No-Fault Punishment*, 40 ALA. L. REV. 1037, 1037 (1989) (emphasizing unpredictability). There is also systematic evidence of growth in awards. A study surveying five large states showed increasingly large awards against businesses, from a total of \$1.1 million from 1968 to 1971 to \$343 million from 1988 to 1991, an 89-fold increase after correcting for inflation. See STEPHEN M. TURNER ET AL., *PUNITIVE DAMAGES EXPLOSION: FACT OR FICTION? 2-4* (Washington Legal Found. Working Paper Series No. 50, 1992). Another recent study found that in San Francisco, California, and Cook County, Illinois, punitive damage awards increased from about \$1 million during the period from 1960-1964, to \$147 million during 1985-1989, and to \$215 million during 1990-1994. See DEBORAH HENSLER & ERIK MOLLER, *TRENDS IN PUNITIVE DAMAGES: PRELIMINARY DATA FROM COOK COUNTY, ILLINOIS AND SAN FRANCISCO, CALIFORNIA* 6 tbl.2 (RAND Inst. for Civil Justice Unrestricted Draft Series No. DRU-1014-ICJ, 1995); see also ERIK MOLLER ET AL., *PUNITIVE DAMAGES IN FINANCIAL INJURY JURY VERDICTS 6-11* (1997) (showing substantial variability and growth in punitive awards).

17. See Stephen Daniels & Joanne Martin, *Myth and Reality in Punitive Damages*, 75 MINN. L. REV. 1, 31-32 (1990).

18. See *id.* at 42 tbl.VI.

19. 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).

20. See *id.* at 147.

21. See, e.g., Product Liability Reform Act of 1997, S. 5, 105th Cong. § 108 (1997) (proposing to cap punitive damage awards); H.R. 956, 104th Cong. § 2 (1995) (proposing to establish uniform legal standards and caps on punitive damage awards); H.R. 955, 104th Cong. § 8(A)-(B) (1995) (proposing to establish

It is not difficult to understand the widespread concern with erratic punitive damage awards. If similarly situated people—plaintiffs and defendants alike—are not treated similarly, erratic awards are unfair. As a matter of fairness, the evidence suggests that some awards are too low, while others are too high. From the standpoint of economic efficiency, unpredictable awards need not be troublesome; perhaps individual awards cannot be calculated in advance, but if people can calculate the expected value of the relevant risks, there should be no efficiency loss. If awards are unpredictable, however, resources are likely to be wasted on that calculation, and as a practical matter, a risk of extremely high awards is likely to produce excessive caution in risk-averse managers and companies.<sup>22</sup> Hence unpredictable awards create both unfairness and (on reasonable assumptions) inefficiency, in a way that may overdeter desirable activity.<sup>23</sup>

Our principal interest here is in identifying some of the sources of unpredictability in jury judgments.<sup>24</sup> On the basis of responses from 899 jury-eligible citizens, we offer a number of findings. Three are particularly important.

(1) People have a remarkably high degree of moral consensus on the degrees of outrage and punishment that are appropriate for punitive damage

---

a clear-and-convincing evidence rule and caps on punitive damage awards).

Academic commentators have also called for legislative and judicial reform. *See, e.g.*, E. Donald Elliott, *Why Punitive Damages Don't Deter Corporate Misconduct Effectively*, 40 ALA. L. REV. 1053, 1072 (1989); *see also* Paul Rubin et al., *BMW v. Gore: Mitigating the Punitive Economics of Punitive Damages*, 1997 SUP. CT. ECON. REV. 179, 184 (arguing that the legal system does not adequately discipline punitive damage awards).

There is no empirical consensus, however, that punitive damages are in fact unpredictable. An important study suggests that the aggregate data show a reasonably orderly pattern, though it acknowledges that the relevant data can be interpreted in different ways. *See* Theodore Eisenberg et al., *The Predictability of Punitive Damages*, 26 J. LEGAL STUD. 623 (1997). *But see* A. Mitchell Polinsky, *Are Punitive Damages Really Insignificant, Predictable, and Rational?*, 26 J. LEGAL STUD. 663 (1997) (responding to the Eisenberg et al. study). In its careful analysis, the Eisenberg et al. study shows that the logarithmic transformation of punitive damage awards yields a distribution that is almost normal. *See* Eisenberg et al., *supra*, at 637, 638 fig.2. It also shows that the log of awards is predicted rather well from a set of objective characteristics of cases in which the awards were made. *See id.* at 644 (using the size of compensatory damages, case categories, plaintiff and defendant status, and locale). Eisenberg et al. thus conclude that the unpredictability of punitive awards has been overstated. *See id.* at 659-60.

Eisenberg et al.'s analyses of real jury awards are not inconsistent with our experimental data. Indeed, we agree with their conclusion that log awards may be normally distributed. But defendants and plaintiffs live in a world of real dollars, not of log dollars. In terms of real dollars, the judgments of our respondents and of the juries examined by Eisenberg et al. are correctly described as erratic and unpredictable, because the severe skewedness creates the possibility of either modest or disastrous losses in identical cases. This produces unfairness and could induce risk aversion even in very large firms.

22. *See generally* KENNETH R. MACCRIMMON & DONALD A. WEHRUNG, *TAKING RISKS: THE MANAGEMENT OF UNCERTAINTY* (1986) (studying responses of business executives to risk).

23. For relevant evidence, see Richard L. Manning, *Products Liability and Prescription Drug Prices in Canada and the United States*, 40 J.L. & ECON. 203, 210-34 (1997), which explains the factors that account for drug price differences; and Rubin et al., *supra* note 21, at 192-96, which discusses the significant market effects of punitive damage awards.

24. We do not contend that we have identified the only sources. *See infra* Section IV.A. There are also questions about the difference between experimental and real world conditions. *See infra* note 121; *see also infra* note 143 and accompanying text.

cases.<sup>25</sup> At least in the products liability cases we offer, this moral consensus, on what might be called outrage and punitive intent, cuts across differences in gender, race, income, age, and education. For example, our study shows that all white, all black, all Hispanic, all female, all male, all poor, all wealthy, all old, and all young juries are likely to come to similar conclusions about how to rank a range of cases.

(2) This consensus fractures when the legal system uses dollars as the vehicle to measure moral outrage. Even when there is a consensus on punitive intent, there is no consensus about how much in the way of dollars is necessary to produce appropriate suffering in a defendant. Under existing law, widely shared and reasonably predictable judgments about punitive intent become highly erratic judgments about appropriate dollar punishment. A basic source of arbitrariness within the existing system of punitive damages (and a problem not limited to the area of punitive damages) is the use of an unbounded dollar scale.

(3) A modest degree of additional arbitrariness is created by the fact that juries have a hard time making appropriate distinctions among cases when they are not comparing them directly. When one case is seen apart from other cases, people show a general tendency to place it toward the midpoint of any bounded scale. It is therefore less likely that sensible discriminations will be made among diverse cases. In producing arbitrary awards, however, this effect is far less important than the difficulty of translating shared moral judgments into dollar values.

The principal purpose of this Article is to elaborate these findings and to use them to develop some policy reforms in the area of punitive damages. Our basic suggestion is that the legal system should enable juries to engage in tasks that they are capable of performing and should not require juries to carry out tasks that they cannot perform well. Juries are likely to produce erratic judgments about dollar amounts; their judgments are likely to be much less erratic when they are asked to rank cases or to assess the degree to which a defendant should be punished on a bounded rating scale. If juries cannot consistently or sensibly “map” their judgments onto an unbounded dollar scale, perhaps the civil justice system should be brought more closely in line with the criminal justice system, in which juries decide questions of liability and judges decide questions of punishment subject to guidelines and constraints.<sup>26</sup> The

---

25. Two qualifications are necessary. First, this conclusion is restricted to the area we investigate, involving personal injury suits. It is an open question whether the moral consensus would operate in areas involving, for example, sexual harassment and discrimination on the basis of race. Second, there is a greater consensus on how to rank the scenarios than on the “absolute” numbers for outrage and punishment.

26. The appropriateness of this conclusion depends partly on the importance of predictability in punitive damage awards. Such awards may be predictable but nonetheless problematic because, for example, they are too high or too low to produce optimal deterrence or because they do not reflect the right moral theory (perhaps one of retribution). Our focus here is on the problem of unpredictability, which is an affront to the rule of law and a particular problem under the Due Process Clause and associated principles. We discuss below the relationship between unpredictability and other possible problems with damage awards. See *infra* Section II.B.



same conclusion might also apply to compensatory awards in the types of cases in which arbitrariness in dollar awards is likely. Alternatively, the current system of civil liability, both punitive and compensatory, might be displaced in some cases by an administrative process seeking to achieve greater coherence and rationality.

The choice of reform depends on an assessment of precisely which defects ought to be corrected and thus on a set of normative judgments about the problems with punitive awards in their current form. We attempt to disaggregate three possible judgments here, partly as thought experiments, and partly to raise some questions about the appropriate domain of populist elements in the legal system.

If the basic problem is simple unpredictability, the legal system might reduce that problem by asking juries not to come up with dollar amounts, but to rank the cases at hand among a preselected set of exemplar cases or by using a bounded scale of numbers rather than an unbounded scale of dollars. A conversion formula, based on previously compiled population-wide data, might be used to generate population-wide judgments about dollar amounts. The result would be a form of *predictable populism*. Through this route, it would be possible to reduce variability and to ensure that jury judgments about appropriate dollar punishments do not reflect the likely unrepresentative views of twelve randomly selected people, but instead those of the population as a whole.

If the basic problem is that people cannot sensibly map their moral judgments onto dollar awards, the legal system should provide a mechanism by which judges or administrators, instead of jurors, can translate the relevant moral judgments into dollar amounts. It is reasonable to question whether ordinary people can know what a given dollar amount would mean for, or do to, the defendant or those in the position of the defendant. On this view, the jury should also rank the case at hand in comparison to preselected cases or come up with a number on a bounded scale. A conversion formula, based not on population-wide data but on expert judgments about what various awards would actually mean or do, would be used to produce rational judgments about dollar amounts. The result would be a form of *technocratic populism*—populist in the sense that normative judgments of the jury would be the foundation of decision, technocratic in the sense that experts would translate those judgments into legal awards.

If the basic problem is that people's moral judgments are not the proper basis for punitive awards, judges might, in some or all contexts, use those moral judgments as one factor to be considered among others, or the legal system might dispense with jury judgments entirely in some or all contexts. If, for example, it is believed that existing social norms are not the appropriate basis for punishment, or if deterrence rather than retribution is the appropriate goal of punitive damages, an expert body might decide on appropriate awards,

or it could offer general guidance to trial court judges. The result would be a form of *bureaucratic rationality*.

We discuss these points in some detail. We also compare these reform proposals with other alternatives, including punitive damage “caps,” simple multipliers, greater judicial control of awards, and monetary schedules building on the workers’ compensation model. We suggest that caps and multipliers have serious problems but that firmer judicial control—on a kind of civil sentencing model—would likely be a desirable and easily administrable step, at least if judges can produce or work from monetary schedules or otherwise attempt to work from comparison cases. A more radical reform, with much promise but also some risk, would involve a shift from juries to administrators operating from a set of guidelines produced by specialists in the areas at hand, subject of course to democratic safeguards.

Both the empirical findings and the policy recommendations have implications well beyond the area of punitive damages. The problem of “mapping” onto a dollar scale arises not only in the setting of punitive damages, but also in damages for pain and suffering, libel actions, sexual harassment cases, intentional infliction of emotional distress, and administrative penalties for injuries that are hard to monetize.<sup>27</sup> From our findings, it is reasonable to infer that the phenomenon of shared judgments but erratic or arbitrary<sup>28</sup> dollar amounts creates a serious and unaddressed problem in many areas of both private and public law.

More generally, our findings raise a fundamental question: How can the normative goals of the legal system be made to mesh with what is, or might be, known about human psychology? This is a large and unanswered question. We attempt to make progress on it here. In the past three decades, a great deal of progress on answering that question has come from the application to legal problems of a certain understanding of psychology embodied by economic theory.<sup>29</sup> Within economics and psychology, but outside law, that understanding has been under sustained attack.<sup>30</sup> Within social science generally, this attack has produced insights that supplement, and sometimes undermine, those versions of economics that have undergirded the economic analysis of law.<sup>31</sup> Our study helps show what might be missing, impractical,

---

27. See W. KIP VISCUSI, *REFORMING PRODUCTS LIABILITY* 102 (1991).

28. They may be arbitrary either in the sense of being erratic or in the sense of being a product of a normatively questionable “anchor” on which the jury has seized. See *infra* Subsection IV.A.2.

29. For a wide range of relevant discussions, see RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* (5th ed. 1997).

30. A good recent collection of essays on the subject is *RESEARCH ON JUDGMENT AND DECISION MAKING* (William M. Goldstein & Robin M. Hogarth eds., 1997).

31. There are some important exceptions. See, e.g., Linda Babcock & George Loewenstein, *Explaining Bargaining Impasse*, 11 *J. ECON. PERSP.* 109 (1997); Mark Kelman et al., *Context-Dependence in Legal Decision Making*, 25 *J. LEGAL STUD.* 287 (1996); Edward J. McCaffery et al., *Framing the Jury: Cognitive Perspectives on Pain and Suffering Awards*, 81 *VA. L. REV.* 1341 (1995).

or wrong in standard economic approaches to punitive damage awards,<sup>32</sup> and bears on a range of additional issues as well.

Part II of this Article outlines existing understandings about punitive damages and shows how the sources of variable judgments are relevant to constitutional issues and to existing theories of deterrence and retribution. Part III outlines our study and its basic conclusions. Part IV presents policy recommendations, most of which are designed to provide a role for community judgments without the noise and arbitrariness that accompany the current system. What are important are not the details of the proposals but their basic goal and direction: to provide juries with an opportunity to do what they are capable of doing relatively well, without requiring them to do what they are bound to do badly. An alternative, of course, is to dispense with juries entirely.

Part V briefly discusses some analogies and implications, dealing above all with the general problem of jury determination of dollar amounts in contexts in which monetization is unfamiliar and difficult. We show how an evaluation of appropriate reforms with respect to compensatory awards calls for an ambitious decision about the role of populist and technocratic ingredients of “compensation” judgments—or, in other words, a decision about what the vexing idea of “compensation” should be taken to mean. We also outline a possible research agenda to see how actual or potential descriptive findings in psychology might bear on the normative goals of the legal system.

## II. DETERRENCE, RETRIBUTION, AND THE CONSTITUTION

### A. *The Goals of Punitive Damages*

The traditional view is that punitive awards serve deterrent and retributive goals. A standard jury instruction says:

In determining whether or not you should award punitive damages, you should bear in mind that the purpose of such an award is to punish the wrongdoer and to deter that wrongdoer from repeating such wrongful acts. In addition, such damages are also designed to serve as a warning to others, and to prevent others from committing such wrongful acts.<sup>33</sup>

---

32. For an important recent economic analysis of punitive damages, see Polinsky & Shavell, *supra* note 7. Polinsky and Shavell argue that juries should decide punitive damage problems by reference to the likelihood that the defendant's acts will go undetected. *See id.* at 873-75. The principal difficulty with this suggestion is that it does not deal with psychological and institutional problems that sharply limit jurors' capacity and willingness to analyze the punitive damage problem in the way the authors suggest. Thus, a reasonable inference from the Polinsky-Shavell analysis is that punitive damages should be assessed by an administrative agency or some other expert body, not by juries or even judges. *See infra* Subsection IV.B.3.

33. RONALD W. EADES, JURY INSTRUCTIONS ON DAMAGES IN TORT ACTIONS 98 (3d ed. 1993).

It is not our purpose to resolve the debate about whether and why a court should award punitive damages.<sup>34</sup> Our modest objective is to outline some of the relevant arguments by way of introduction. These arguments provide the foundation for the constitutional law governing punitive damage awards, and they play a large role in our empirical study.

## 1. *Deterrence*

### a. *Conventional Arguments*

With respect to the goal of deterrence, there is a simple and standard economic argument for punitive damages: Compensatory damages work well for deterrence if and only if it is easy to identify the injurer, bring suit, and collect full damages. Under these conditions, the wrongdoer faces liability for the full social costs of the wrong. Thus there is no need for additional damages if the probability of detection and successful suit for compensation is 100%. But sometimes it is difficult to identify the injurer, perhaps because the tort has occurred surreptitiously. In such cases, compensatory damages will not provide adequate deterrence, since wrongdoers will be able to continue to engage in conduct where the social costs exceed the social benefits. Punitive damages are necessary to pick up the slack for undetected wrongdoers.<sup>35</sup>

Under this view, the goal of punitive damages is to ensure that the award of compensatory damages is supplemented by an amount sufficient to cause wrongdoers to internalize the costs of their actions. The simplest conclusion is that total damages should be the harm multiplied by the reciprocal of the probability that the defendant will be found liable when he should be found liable; punitive damages would then consist of the excess of total damages over compensatory damages.<sup>36</sup> We will suggest some institutional and psychological reasons to doubt that jurors can or will make judgments of this kind; let us continue with the simple account for now.

With some torts especially—medical malpractice is an obvious example—many defendants may be able to escape compensatory damages. When the defendant has been able to conceal his identity, or otherwise to escape being sued, there is therefore good reason to award punitive damages. Interestingly, this analysis supports a relatively high punitive damage award in

---

34. Relevant discussions include Robert D. Cooter, *Economic Analysis of Punitive Damages*, 56 S. CAL. L. REV. 79 (1982); Robert D. Cooter, *Punitive Damages for Deterrence: When and How Much?*, 40 ALA. L. REV. 1143 (1989); Dorsey D. Ellis, *Fairness and Efficiency in the Law of Punitive Damages*, 56 S. CAL. L. REV. 1 (1982); Galanter & Luban, *supra* note 8; Owen, *supra* note 8; and Polinsky & Shavell, *supra* note 7.

35. See LANDES & POSNER, *supra* note 7, at 160-63; Polinsky & Shavell, *supra* note 7, at 874.

36. See Polinsky & Shavell, *supra* note 7, at 874-75.

the *BMW* case,<sup>37</sup> because it is not easy to detect fraudulent repainting of cars. But in cases in which the probability of detection and suit is extremely high, punitive damages make far less sense. This may be true, for example, of some assaults or homicides.

From the standpoint of optimal deterrence, there is a second possible reason to award punitive damages. There may be cases in which a social judgment (not itself based on purely economic factors) is made that certain subjective gains ought not to be allowed to count in a global utility calculus.<sup>38</sup> If someone has gained utility from murder, rape, assault, or sexual harassment, it might be thought (reasonably enough) that such a gain should not be treated as a benefit to be weighed against the costs of such activity. Punitive damages are necessary to offset the utility gain that is judged illicit.<sup>39</sup> Compensatory damages are inadequate in such cases because they do not have this effect; they require the defendant to internalize the victim's costs without also eliminating the hedonic effects of the defendant's benefits.

A third possible rationale for punitive damages is that sometimes compensatory damages may be lower than they should be, particularly where it is especially costly for courts to calculate compensatory damages. If compensatory awards are systematically low, and if calculation costs are high, the amount awarded for punitive damages may move the legal system closer to optimal deterrence.<sup>40</sup> Of course, this argument has a degree of speculativeness and crudeness, because it is unlikely that punitive damages can be calibrated with sufficient precision to make up for the shortfall in compensatory damages, and because compensatory awards might in fact be at the right level or even too high in the aggregate. Nevertheless, punitive damage awards bring the incentives of wrongdoers closer to where they should be.

On the deterrence view, intentional torts, i.e. those involving deliberate infliction of an injury, may provide particularly appropriate cases for punitive damage awards. In such cases, socially illicit gains are most likely to be involved, and intentional torts frequently (though not always) involve a form of concealment. But an implication of the deterrence view is that punitive damages generally do not make sense for highly visible torts where the probability of detection and compensation is extremely high. In such cases, compensatory damages are all that is required.<sup>41</sup>

---

37. *BMW of N. Am., Inc. v. Gore*, 116 S. Ct. 1589 (1996) (invalidating a punitive damage award for a repainted car).

38. See, e.g., John C. Harsanyi, *Morality and the Theory of Rational Behavior*, in *UTILITARIANISM AND BEYOND* 39, 56 (Amartya Sen & Bernard Williams eds., 1982) (proposing the exclusion of sadistic or malicious preferences from utility calculations).

39. See Polinsky & Shavell, *supra* note 7, at 918.

40. See LANDES & POSNER, *supra* note 7, at 161.

41. See *id.* at 160; Polinsky & Shavell, *supra* note 7, at 878.

b. *Puzzles and Problems with the Deterrence Justification*

Even from the economic point of view, these deterrence-based arguments for punitive damages raise some serious puzzles. An important question, not yet taken up in the economic literature, is the relationship between this rationale for punitive damages and the existence of other regulatory controls. Both administrative and criminal law can be defended as efforts to compensate for private underenforcement of law;<sup>42</sup> if the tort system also attempts to pick up the slack via punitive damages, overdeterrence may well result. A system containing compensatory and punitive damages may be both necessary and sufficient taken by itself, but if it is complemented by administrative and criminal law, it is likely to become incoherent.

Even without administrative and criminal law, efforts to impose punitive damages may misfire because of the inevitable confusion or the cognitive and motivational errors of jurors and judges. There is thus a question whether real world institutions can reliably engage in the enterprise of seeking to obtain optimal deterrence. That enterprise is very costly, and the costs of decision may argue in favor of some other system for calculating punitive damages. Perhaps juries will not understand the inquiry into optimal deterrence; perhaps it will be too costly to give them the relevant understanding; perhaps they will refuse to undertake that inquiry even if they are asked to do so. We return to these issues below.<sup>43</sup>

If juries are not thinking in economic terms, they may award large punitive damages in a way that deters desirable activity (or award small punitive damages in a way that produces insufficient deterrence). With large awards, activity that produces benefits as well as costs may be stopped or significantly reduced—with possible adverse effects on safety and health itself.<sup>44</sup> It is for this reason that the strongest case for punitive damages involves intentional torts and that the weakest case involves negligence. In cases involving negligence, punitive damages would be likely to reduce desirable activity. This is much less likely for intentional wrongdoing, where overdeterrence of

---

42. See, e.g., POSNER, *supra* note 29, at 240. For a more general discussion, see Edward Rubin, *Punitive Damages: Reconceptualizing the Runcible Remedies of Common Law*, 1998 WIS. L. REV. (forthcoming May 1998).

43. See *infra* Subsection IV.A.4.

44. Thus, there is a possibility here of health-health tradeoffs, which occur when a strategy designed to reduce health risks actually increases such risks. For example, some punitive damage awards against pharmaceutical companies may actually have adverse health effects, if they serve to increase prices. For a recent discussion raising this issue in the context of products liability risk and international prescription drug price differentials, see Manning, *supra* note 23, at 210-34. For general discussions of the concept of health-health tradeoffs, see JOHN D. GRAHAM & JONATHAN BAERT WEINER, *RISK VERSUS RISK: TRADEOFFS IN PROTECTING HEALTH AND THE ENVIRONMENT* (1996), which examines the phenomenon of making risk tradeoffs when formulating policy regarding health and the environment; and Cass R. Sunstein, *Health-Health Tradeoffs*, 63 U. CHI. L. REV. 1533 (1996), which argues that administrative agencies should have the authority to make health-health tradeoffs when formulating regulations. Evidence of such effects in the context of punitive damages can be found in Rubin et al., *supra* note 21, at 192-96.

desirable activity is far less likely to be at issue. Reckless behavior is of course an intermediate case.

For those interested in optimal deterrence, a particularly important dispute is whether, on economic grounds, the wealth or income of the defendant should matter.<sup>45</sup> As we will explain, our findings suggest that the defendant's wealth is likely to be an important variable in actual damage awards, though it plays little or no role in people's judgments about outrageousness or appropriate punishment on a bounded numerical scale.<sup>46</sup> On a conventional view about optimal deterrence, however, wealth and income are irrelevant. Properly calculated by reference to the probability of punishment and the need to counteract illicit gains, a punitive damages award should encourage a defendant to engage in optimal behavior, whatever its wealth.<sup>47</sup> But there is a question whether this is realistic; we take up this issue below.<sup>48</sup>

## 2. Retribution

Thus far we have spoken in terms of deterrence, but punitive damages also have an expressive or retributive purpose. They are designed to punish as well as to deter. Juries believe that such awards express the community's outrage at certain forms of behavior, and judges' instructions encourage juries to think in precisely these terms. In fact, empirical evidence—including that provided here<sup>49</sup>—suggests that juries are not attempting to promote optimal deterrence but instead to punish wrongdoing with, at most, a signal designed to ensure that certain misconduct will not happen again.

Regrettably, the legal culture lacks a full normative account of the relationship between retributive goals and punitive damages. Those who emphasize retribution are seeking to ensure that the punishment reflects the community's outrage about certain acts. The retributive idea would probably focus on two principal factors: the defendant's state of mind and the degree of

---

45. A number of courts and commentators have argued that wealth should generally be irrelevant. *See, e.g., Kemezy v. Peters*, 79 F.3d 33, 35 (7th Cir. 1996) ("What is striking about the purposes that are served by the awarding of punitive damages is that none of them depends critically on proof that the defendant's income exceeds some certain level."); Kenneth S. Abraham & John C. Jeffries, Jr., *Punitive Damages and the Rule of Law: The Role of Defendant's Wealth*, 18 J. LEGAL STUD. 415, 415 (1989) (arguing that "the defendant's wealth is irrelevant to the goal of deterring socially undesirable conduct"); Ellis, *supra* note 34, at 63 (arguing that the use of wealth as a factor "invites juries to engage in wealth redistribution and exacerbates the perverse incentives already created by uncertain standards of punitive damage liability"); Polinsky & Shavell, *supra* note 7. For an argument the other way, see Jennifer H. Arlen, *Should Defendants' Wealth Matter?*, 21 J. LEGAL STUD. 413 (1992), which discusses the deterrence of risk-averse people.

46. *See infra* notes 133-135 and accompanying text; cf. Robert J. MacCoun, *Differential Treatment of Corporate Defendants by Juries: An Examination of the "Deep-Pockets" Hypothesis*, 30 L. & SOC'Y REV. 121, 133-39 (1996) (finding effects on jury awards from corporate identity and commercial activity, but not from wealth per se).

47. *See* Polinsky & Shavell, *supra* note 7, at 911.

48. *See infra* Subsection IV.B.3.

49. *See infra* Part III.

harm actually caused or likely to be caused by the defendant's behavior. The first point is especially important. Civil or criminal punishment may attempt to capture the intuition that certain intentional states of mind are especially blameworthy because they reflect abhorrent and false views about the moral worth of persons.<sup>50</sup> When one person murders another, or acts with gross negligence that produces another's death, the wrongdoer reveals a belief that some people do not matter very much.<sup>51</sup> In this sense, retribution properly understood embodies a principle of moral equality.<sup>52</sup>

The award of punitive damages can also be taken to have an expressive function: It expresses the community's outrage at certain forms of conduct, in a way that is intended both to reflect and to entrench the relevant social norms. This idea connects the award of punitive damages with their historical origins in affronts to the honor of the victims.<sup>53</sup> This understanding helps explain the provision of punitive damage awards for sexual harassment and other violations of civil rights statutes.<sup>54</sup> The relevant community now believes that these kinds of illegality reflect an especially blameworthy state of mind, and hence that punitive damages are a necessary supplement to other sanctions. The retributive idea thus suggests that the most serious cases for punitive damage awards involve harms that are both grave in degree and affirmatively desired by the defendant.

We connect these factors to the psychology of punitive damage awards in the discussion below.<sup>55</sup> The central points are that people appear to be intuitive retributivists, and if retribution is the goal of punitive damage awards, it is important to ensure that juries are asked questions that allow them to express, in a rational and coherent manner, their sense of outrage and their judgments about appropriate punishment. The question then becomes whether dollar amounts are a sensible register of that sense and those judgments.

## B. *The Constitution and Punitive Damages*

### 1. *Substantive Versus Procedural Interpretations of Unconstitutional Punishment*

What does the Constitution have to say about the award of punitive damages? Because punitive damage awards have been so controversial, and

---

50. See Hampton, *supra* note 8, at 124-38.

51. See *id.*; see also Galanter & Luban, *supra* note 8, at 1432-38 (connecting punitive awards with retributive goals and community outrage); Jean Hampton, *Correcting Harms Versus Righting Wrongs: The Goal of Retribution*, 39 UCLA L. REV. 1659, 1687-89 (1992) (discussing punitive damage awards).

52. Of course, retributivists need not be moral egalitarians; we could imagine retributivists who thought that harms against some persons deserve less punishment than harms against other persons.

53. See Ellis, *supra* note 34, at 14-15.

54. See, e.g., 42 U.S.C. § 1981a(b)(1) (1994).

55. See *infra* Section IV.B.



because the Supreme Court has been effectively forced to evaluate a range of punitive damage judgments in scattered decisions, this question has become one of the more complex and pressing in modern constitutional law. There are three plausible answers: (1) Punitive damages awards are always constitutional; (2) such awards are unconstitutional, as violations of the Due Process Clause of the Fifth Amendment and the Fourteenth Amendment in its substantive dimension, when they are grossly excessive; (3) such awards are unconstitutional, as violations of the Due Process Clause in its procedural dimension, when they are too likely to be arbitrary—because, for example, they are unaccompanied by procedures that sufficiently confine the discretion of the jury. Propositions (2) and (3), of course, are not mutually exclusive. The Constitution might prohibit both excessive and arbitrary awards, or awards that are very high but not adequately subject to the rule of law. Our findings here are directed principally toward proposition (3) and to the general questions of whether, why, and in what sense punitive damage awards are likely to be unpredictable and arbitrary. The constitutional judgment overlaps with the more mundane work of district judges and appellate courts, which set aside many punitive awards as arbitrary because they are out of step with the facts of the case or with comparison cases.<sup>56</sup>

Recently, most notably in *BMW of North America, Inc. v. Gore*,<sup>57</sup> a majority of the Supreme Court has converged on proposition (2), though propositions (1) and (3) have also received significant support, and (3) is highly likely to be a prime area for future contestation. In fact, the *BMW* majority accompanied its endorsement of a version of (2) with a statement that a major problem with excessive awards is that they fail to provide sufficient predictability.<sup>58</sup> Hence a majority of recent Justices—most prominently Justices O'Connor<sup>59</sup> and Breyer,<sup>60</sup> but at different times also Justices Souter, White,<sup>61</sup>

---

56. A study of products liability punitive damages cases between 1965 and 1990 shows 355 awards. See Michael Rustad, *In Defense of Punitive Damages in Products Liability*, 78 IOWA L. REV. 1, 38 (1992). Of the non-asbestos-related cases, roughly 40% ultimately settled, a quarter of punitive damages were affirmed on appeal, and nearly a third of verdicts were ultimately reversed or remitted. See *id.* at 54-55. A study by William Landes and Richard Posner of products liability cases found that punitive damages in state cases were upheld in fewer than 2% of 119 cases, a percentage lower than in federal cases studied. See William M. Landes & Richard A. Posner, *New Light on Punitive Damages*, REGULATION, Oct. 1986, at 33, 35-36; see also LANDES & POSNER, *supra* note 7, at 302-07 (summarizing and updating this study). Notably, however, products liability suits do not produce the bulk of punitive damage awards. Apparently, about 47% are in business contract cases and 36% in intentional tort cases. See ERIK MOLLER, TRENDS IN CIVIL JURY VERDICTS SINCE 1985, at 34 (1996).

57. 116 S. Ct. 1589 (1996).

58. See *id.* at 1604; see also *id.* at 1605-07 (Breyer, J., concurring); Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 42-43 (1991) (O'Connor, J., dissenting).

59. See *Browning-Ferris Indus., Inc. v. Kelco Disposal, Inc.*, 492 U.S. 257, 282-83 (1989) (O'Connor, J., concurring in part and dissenting in part).

60. See *BMW*, 116 S. Ct. at 1605-07 (Breyer, J., concurring).

61. Justices Souter and White joined the relevant parts of Justice O'Connor's dissenting opinion in *TXO Production Corp. v. Alliance Resources Corp.*, 509 U.S. 443, 472-73 (1993).

Brennan,<sup>62</sup> Marshall,<sup>63</sup> Stevens,<sup>64</sup> Kennedy,<sup>65</sup> and Blackmun<sup>66</sup>—have argued that the Due Process Clause requires constraints on jury discretion that will provide fair notice to potential defendants and limit the role of arbitrary or irrelevant factors. The Court has not given much of an account of why juries might be arbitrary or unpredictable; our findings offer the elements of such an account—one that attempts to specify and model some of the concerns voiced by Justices O'Connor and Breyer.

In a series of cases in the 1990s the Court refused to set aside punitive damage awards as inconsistent with the Excessive Fines Clause of the Eighth Amendment<sup>67</sup> or the Due Process Clause.<sup>68</sup> But it left open the possibility that in an extreme case, an award would be constitutionally unacceptable under the Due Process Clause. The punitive damages cases have thus refused to endorse the rule, proposed most insistently by Justice Scalia,<sup>69</sup> that the Constitution imposes no constraints on what juries may do in this context. Instead, the Court has left open the possibility of a due process limitation on excessive awards. The Court has stressed the possibility of a successful claim on the basis of substantive rather than procedural due process insofar as it has emphasized that the problem would arise if the relevant award was unjustifiably large or “excessive”—rather than if the procedure that produced those awards was unreliable. There is a clear parallel here to due process arguments against capital punishment; one argument is that the death penalty is sometimes or always excessive, while another is that judgments of death are unconstitutional if produced by insufficiently reliable procedures. After seriously considering both routes, the Court eventually converged on the procedural option with respect to the death penalty, and it is possible that this will happen with punitive damage awards as well.

## 2. *Leading Cases*

For constitutional purposes, the key case is *BMW of North America, Inc. v. Gore*.<sup>70</sup> In that case, Dr. Ira Gore sought punitive damages because he had not been informed that his new BMW was actually repainted. The jury granted an award of punitive damages of \$4 million, an award that seemed well out of line with the \$4000 compensatory damages award. (On appeal to the Alabama

---

62. See *Browning-Ferris*, 492 U.S. at 280-82 (Brennan, J., concurring).

63. See *id.* (Brennan, J., concurring, joined by Marshall, J.).

64. See *BMW*, 116 S. Ct. at 1598-1604 (Stevens, J., concurring).

65. See *TXO*, 509 U.S. at 466-67 (Kennedy, J., concurring in part and concurring in the judgment); *Pacific Mut. Life Ins. Co. v. Haslip*, 499 U.S. 1, 40-42 (1991) (Kennedy, J., concurring in the judgment).

66. See *Haslip*, 499 U.S. at 18-24 (Blackmun, J.).

67. See *Browning-Ferris*, 492 U.S. at 262-64.

68. See *TXO*, 509 U.S. at 462 (plurality opinion); *Haslip*, 499 U.S. at 19-24.

69. See *BMW*, 116 S. Ct. at 1610-12 (Scalia, J., dissenting).

70. 116 S. Ct. 1589 (1996).

Supreme Court, the punitive award was reduced to \$2 million.)<sup>71</sup> Presented with this disparity, the Court ruled for the first time that an award of punitive damages violated the Due Process Clause. The Justices divided, however, on the reason for the violation. The opinion for the five-member majority spoke in terms of excessiveness and hence substantive due process,<sup>72</sup> but included subthemes of federalism<sup>73</sup> and fair notice.<sup>74</sup> Justice Breyer joined the majority, but his concurring opinion was procedurally oriented.<sup>75</sup> The four dissenting Justices seemed to say that no punitive damage award could ever violate the Due Process Clause.<sup>76</sup> Part of the impetus for the dissenters' views may have been that such a bright-line rule would impose good incentives on legislatures to address the punitive damages problem.<sup>77</sup>

In finding the award grossly excessive, the Court indicated that a state may not impose sanctions on law-violators with the goal of changing the violators' lawful conduct in other states. The Court made clear that a state may not attempt to change a company's policies in other states if those policies are not unlawful in those other states,<sup>78</sup> and though it left some ambiguity, it suggested that a state may impose punitive damages only to protect its own consumers and its own economy.<sup>79</sup> In a crucial passage, the Court also said that the Due Process Clause requires not merely fair notice of what is criminal, but also fair notice "of the severity of the penalty that a State may impose."<sup>80</sup> In finding that BMW did not receive adequate notice of the magnitude of the sanction that Alabama might impose, the Court referred to three "guideposts":<sup>81</sup> the degree of reprehensibility of the unlawful behavior; the relation between the harm or potential harm suffered on the one hand and the punitive damages award on the other; and the difference between the punitive award and the available penalties for comparable misconduct. Taking these guideposts together, the Court found the award unconstitutional because it was "grossly excessive."<sup>82</sup>

---

71. See *BMW of N. Am., Inc. v. Gore*, 646 So.2d 619 (Ala. 1994), *rev'd*, 116 S. Ct. 1589 (1996).

72. See *BMW*, 116 S. Ct. at 1598-604.

73. See *id.* at 1595-98.

74. See *id.* at 1598 & n.22.

75. See *id.* at 1604-09 (Breyer, J., concurring).

76. See *id.* at 1610-14 (Scalia, J., joined by Thomas, J., dissenting) (arguing that the Constitution does not make the concern regarding excessive punitive awards "any of our business"); *id.* at 1614-18 (Ginsburg, J., joined by Rehnquist, C.J., dissenting) (arguing that punitive awards are "an area dominantly of state concern").

77. Cf. *id.* at 1617-20 (Ginsburg, J., dissenting) (discussing state responses to the punitive damages problem). Compare this stance with the idea of a "penalty default" in the law of contracts and statutory construction. See Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 *YALE L.J.* 87, 97 (1989) (discussing penalty defaults for interpretation of contracts); *id.* at 97-98 (discussing penalty defaults for interpretation of statutes).

78. See *BMW*, 116 S. Ct. at 1597.

79. See *id.* at 1596-98.

80. *Id.* at 1598.

81. *Id.*

82. *Id.*

The first factor, in the Court's view, seeks to ensure some proportionality so that damages reflect the extent of the offense.<sup>83</sup> Here retributive goals appear central. In the Court's view, this was a serious problem with the *BMW* jury's award because no special aggravating considerations were present. The injury was purely economic. There was no effect on the car's performance, safety, or even appearance for a significant period. The failure to disclose the relevant material was very plausibly a wrong, especially insofar as it formed part of a national pattern, but it was not a matter of "deliberate false statements," the "concealment of evidence of improper motive," or "egregiously improper conduct."<sup>84</sup>

In discussing the second factor, the Court said that the ratio of punitive damages to actual (or potential) compensatory damages was especially disproportionate: over 500 to 1.<sup>85</sup> Importantly, however, the Court indicated that the ratio alone was not decisive. If an especially egregious act produced a small amount of economic damage, high punitive damages could be legitimate. They could also be legitimate where "the injury is hard to detect or the monetary value of noneconomic harm might have been difficult to determine."<sup>86</sup> But a 500 to 1 ratio in this case was "breathtaking."<sup>87</sup>

With respect to the third factor, the civil and criminal penalties that could be imposed for comparable misconduct were far more limited—involving, for example, a maximum civil penalty for deceptive trade practices of \$2000.<sup>88</sup> Thus the punitive damage award was quite inconsistent with legislative judgments about the relevant conduct as seen in other areas of the law. This comparative inquiry serves as a means of checking jury determinations against the assessments of democratically elected legislatures.

In a concurring opinion joined by Justices O'Connor and Souter, Justice Breyer pressed some different points, drawing on the basic principles of constrained discretion, fair notice, and uniformity of treatment.<sup>89</sup> For Justice Breyer, the most serious problem in the case was not sheer excessiveness but rather the absence of legal standards that could reduce jury discretion and hence caprice. Here the relevant standards "are vague and open-ended to the point where they risk arbitrary results."<sup>90</sup> The Alabama jury operated under no statute with standards distinguishing among permissible punitive damage awards. In *Green Oil Co. v. Hornsby*,<sup>91</sup> decided in 1989, the Alabama

---

83. *See id.* at 1599.

84. *Id.* at 1601.

85. *See id.* at 1602 (comparing the ratio here with those in *Pacific Mutual Life Insurance Co. v. Haslip*, 499 U.S. 1, 7 & n.2 (1991), which involved a 4-to-1 ratio, and *TXO Production Corp. v. Alliance Resources Corp.*, 509 U.S. 443, 453 (1993) (plurality opinion), which involved a 500-to-1 ratio).

86. *Id.*

87. *Id.* at 1603.

88. *See id.*

89. *See id.* at 1605 (Breyer, J., concurring).

90. *Id.*

91. 539 So. 2d 218 (Ala. 1989).

Supreme Court had set out seven factors to be used by state appellate courts to constrain punitive damages awards. These factors were whether there was a reasonable relation to actual and likely harm; the degree of reprehensibility; the amount necessary to remove the profit gained by the unlawful act; the financial position of the defendant; the costs of litigation and the need to create incentives for private litigants; the imposition, or not, of criminal sanctions on the defendant; and the existence of other civil penalties as a mitigating factor.<sup>92</sup> In that case, however, the seven factors were not applied in a way that provided any actual constraint, nor did state courts subsequently make any effort to discipline those factors in such a way as to generate a legally constraining standard. As a permissible ground for punitive awards, Justice Breyer referred to the possibility of permitting “juries to calculate punitive damages by making a rough estimate of global harm, dividing that estimate by a similarly rough estimate of the number of successful lawsuits that would likely be brought, and adding generous attorneys fees and other costs.”<sup>93</sup> Here, there was no evidence that the Alabama Supreme Court applied “any ‘economic’ theory” to explain the \$2 million recovery.<sup>94</sup> Nor was there a community understanding or historic practice that would provide background standards exemplified in that recovery.<sup>95</sup> The general problem lay in the violation of the rule of law: “The upshot is that the rules that purport to channel discretion in this kind of case, here did not do so in fact.”<sup>96</sup>

Thus, Justice Breyer’s opinion can be understood as connecting the outcome in *BMW* with void-for-vagueness cases<sup>97</sup> and the constitutional attack on the death penalty in *Furman v. Georgia*.<sup>98</sup> The central problem lies in unconstrained discretion. For Justice Breyer, the outcome in the *BMW* case is best understood by reference to problems other than excessiveness. We will refer to Justice Breyer’s approach in a number of places below. The problems identified in this Article connect directly with his procedural concerns, but they also suggest a distinctive source of discretion, one very different from that emphasized in his opinion.<sup>99</sup>

---

92. *See id.* at 223-24.

93. *BMW*, 116 S. Ct. at 1607 (Breyer, J., concurring).

94. *See id.* at 1608.

95. *See id.*

96. *Id.* at 1609.

97. *See, e.g.,* *Papachristou v. City of Jacksonville*, 405 U.S. 156 (1972).

98. 408 U.S. 238, 240 (1972) (per curiam) (striking down capital punishment unaccompanied by sufficient constraints on jury discretion).

99. Two other cases should be briefly discussed. In *Pacific Mutual Life Insurance Co. v. Haslip*, 499 U.S. 1 (1991), Haslip complained of the lapsing of his health insurance policy as a result of misconduct by Ruffin, an agent for Pacific Mutual as well as for another, unaffiliated insurance company. Ruffin had misappropriated premiums issued by Haslip’s employer for payment to the other insurer. *See id.* at 5-6. The trial court instructed the jury that it could award punitive damages against Pacific Mutual if it found liability for fraud. *See id.* at 6. The jury awarded a general verdict of \$1,040,000, with a likely division of \$200,000 for compensatory damages and \$840,000 for punitive damages. *See id.* at 7 & n.2. The Court held, in a way that bears a good deal on our study, that the award was acceptable. The instructions referred to deterrent and redistributive goals, pointed to the character and degree of the wrong, and excluded

### 3. *The Need for an Understanding of Jury Variability To Inform Constraints on Punitive Damages*

After *BMW*, and the unruly precedents on which it is based, the law governing constitutional constraints on punitive damage awards is in a state of considerable uncertainty and flux.<sup>100</sup> It is clear that due process questions are raised by any awards that are unaccompanied by limitations on jury discretion and that exhibit striking ratios between punitive damages and compensatory damages (of, say, 10 to 1 or more). It is equally clear that striking ratios are not (and should not be) decisive, that a jury is entitled to consider the wealth of the defendant, and that a plaintiff might be able to eliminate constitutional doubts by showing that the potential harm was higher than the actual harm, that the defendant engaged in a long course of misconduct, or that the defendant's state of mind was especially blameworthy. In *Honda Motor Co. v. Oberg*,<sup>101</sup> the Court held that the Due Process Clause requires judicial review of the size of punitive damage awards to ensure that they are not arbitrary or excessive.<sup>102</sup> The Court said that a state that failed to provide judicial review violated procedural due process, because it risked a "lawless, biased, or arbitrary" result.<sup>103</sup> Thus *Honda Motor* reflects an unambiguous constitutional concern with judgments that are not merely excessive but arbitrary as well. This case joins Justice Breyer's concurring opinion in *BMW*, along with the separately expressed concerns about arbitrary awards from no fewer than nine other recent Justices (with a total of six sitting on the current

---

evidence of Pacific Mutual's wealth. *See id.* at 19-20. Furthermore, the Supreme Court of Alabama had established post-trial procedures for scrutinizing punitive damage awards, *see id.* at 20, and had provided additional checks by examining whether there was a reasonable relation among the award, the actual and likely harm, the degree and duration of the defendant's conduct, the defendant's state of mind, any concealment or past similar conduct by the defendant, the profitability of the conduct to the defendant, and other factors, *see id.* at 20-22.

The Court was badly divided over the damage award in *TXO Production Corp. v. Alliance Resources Corp.*, 509 U.S. 443 (1993) (plurality opinion). TXO had been held liable for \$19,000 in actual damages as a result of having slandered Alliance's title; the jury also awarded \$10 million in punitive damages. *See id.* at 446. The Court upheld the award. *See id.* at 453. Justice Stevens, joined by Chief Justice Rehnquist and Justice Blackmun, wrote the plurality opinion, holding that the jury may properly "tak[e] account of the potential harm that might result from the defendant's conduct," *id.* at 460, and thus that the TXO award was not excessive merely because of "the dramatic disparity between the actual damages and the punitive award," *id.* at 462. If TXO had succeeded, it could have produced a multimillion dollar reduction in its own royalty obligations to Alliance, and Alliance could have suffered a multimillion dollar loss. *See id.* In any case, TXO's pattern of behavior threatened millions of dollars in losses to others, so there was no grotesque disparity between punitive damages and threatened harm. *See id.* In a lengthy dissenting opinion, Justice O'Connor, mostly joined by Justices White and Souter, complained about the procedures underlying the \$10 million award, which, in her view, raised a serious risk of arbitrariness. *See id.* at 474-75 (O'Connor, J., dissenting).

100. *See, e.g.*, *Kim v. Nash Finch Co.*, 123 F.3d 1046, 1067-68 (8th Cir. 1997) (indicating heightened constitutional concerns after *BMW*); *Mathie v. Fries*, 121 F.3d 808, 816-17 (2d Cir. 1997) (same).

101. 512 U.S. 415 (1994).

102. *See id.* at 432.

103. *Id.* at 433.

Court) to show that arbitrary and unpredictable awards raise serious concerns under the Due Process Clause.<sup>104</sup>

But these developments leave many open questions. It is hardly certain that the Court will embark on the project of creating a detailed form of "constitutional common law" to control punitive damages. Yet *BMW* practically forces lower courts to begin to do exactly that, and constitutional doctrine establishing such controls on punitive damage awards is now emerging.<sup>105</sup>

What the Court lacks, and what bears on constitutional controls, is an understanding of the source of jury variability. Where jurors' judgments vary widely, what is the reason? Are jurors reacting to fine-grained distinctions among particular cases? What is the source of unpredictable or erratic judgments? What strategies might work, or fail to work, in counteracting the problem? These are the questions on which we will try to make some progress here.

The answers do not bear only on constitutional law. As a practical matter, nonconstitutional law is far more important as a means of controlling punitive damage awards. Appellate courts must often decide whether to reduce punitive damage awards as unreasonable, sometimes with constitutional pressure in the background. In fact, a significant number of punitive awards do not survive appellate review.<sup>106</sup> To decide when to reduce awards, appellate judges need some understanding of what makes awards unreasonable, and hence an understanding of the sources of variability. Trial courts are in a similar position, both in offering instructions and in reducing awards after they have been made. Even when the Due Process Clause is not at issue, both trial courts and appellate courts will struggle with questions very much like those raised in *BMW* as they continue to devise principles by which to constrain awards. The governing principles have yet to be well settled, and clear standards have yet to emerge. Indeed, it is not at all clear why the legal system should not generally contain a mechanism for increasing punitive damage awards when a particular jury has imposed an unduly low award.<sup>107</sup> If variable and erratic

---

104. See *supra* text accompanying notes 58-66.

105. See, e.g., *Kimzey v. Wal-Mart Stores, Inc.*, 107 F.3d 568, 576-78 (8th Cir. 1997); *Lee v. Edwards*, 101 F.3d 805, 809-12 (2d Cir. 1996); *Stafford v. Puro*, 63 F.3d 1436, 1444-45 (7th Cir. 1995); *Allahar v. Zahora*, 59 F.3d 693, 696-97 (7th Cir. 1995); *Klein v. Grynberg*, 44 F.3d 1497, 1504-05 (10th Cir. 1995); *King v. Macri*, 993 F.2d 294, 298-99 (2d Cir. 1993); *Ross v. Black & Decker, Inc.*, 977 F.2d 1178, 1189-90 (7th Cir. 1992); *Vasbinder v. Scott*, 976 F.2d 118, 121-22 (2d Cir. 1992); *Michelson v. Hamada*, 36 Cal. Rptr. 2d 343, 356-59 (Ct. App. 1994); *Baume v. 212 E. 10 N.Y. Bar Ltd.*, 634 N.Y.S.2d 478, 480 (App. Div. 1995); *Parkin v. Cornell Univ. Inc.*, 581 N.Y.S.2d 914, 916-17 (App. Div. 1992).

106. See LANDES & POSNER, *supra* note 7, at 302-07. But see David Baldus et al., *Improving Judicial Oversight of Jury Damages Assessments: A Proposal for the Comparative Additur/Remittitur Review of Awards for Nonpecuniary Harms and Punitive Damages*, 80 IOWA L. REV. 1109, 1120 n.21 (1995) (claiming that most awards do, in fact, survive appellate review).

107. In some states there is a procedure for "additur," see Baldus et al., *supra* note 106, at 1119-20, but this procedure is used infrequently in punitive damages cases, see *id.*, and not at all in the federal courts, where the Seventh Amendment right to a jury trial bans the use of additur, see *Dimick v. Schiedt*,

judgments are the problem, the legal system should correct awards that are both unreasonably low and unreasonably high, for both are likely to occur.<sup>108</sup>

Of course, policy issues surrounding punitive damages are under active consideration in many states as well as at the national level.<sup>109</sup> States are considering caps in the form of ratios or flat dollar limitations; many states are attempting to require bifurcated trials in which judges rather than juries determine the level of punitive damages.<sup>110</sup> These are somewhat crude and categorical efforts to respond to the general fear that punitive damages are "out of control."<sup>111</sup> An understanding of the source of variability would lead toward more finely tuned remedies, or at least toward a fuller sense of why the crude alternatives would be acceptable. It is this fuller sense that our study is designed to provide.

### III. WHY PUNITIVE DAMAGE AWARDS ARE ERRATIC

We designed a study to examine hypotheses about three topics: (1) the psychology of the sequence of judgments and attitudes that produce individual judgments about punitive awards in particular cases; (2) the sources of variability in these judgments; and (3) the implications of these findings for the unpredictability of jury awards and for possible reforms in the tasks assigned to juries.

---

293 U.S. 474 (1935).

108. The best defense of the current system—which allows judges to reduce or overturn awards that are too high, but not to increase awards that are too low—is that populist enthusiasms might make juries too generous to plaintiffs and too punitive to defendants. There is no fear, apparently, that pro-defendant sentiments will lead in the opposite direction. Our study at least suggests the possibility of problems in both directions, although it particularly provides evidence of "skewing" in the form of high rather than low awards.

109. Three states do not allow punitive damages: Nebraska, New Hampshire, and Washington. A number of states have special restrictions. See ROBERT G. SCHLOERB ET AL., PUNITIVE DAMAGES: A GUIDE TO THE INSURABILITY OF PUNITIVE DAMAGES IN THE UNITED STATES AND ITS TERRITORIES 18-26 (1988). Federal proposals include a cap on punitive damages in some civil suits. See, e.g., Product Liability Reform Act of 1997, S. 5, 105th Cong. § 108 (1997).

110. State law reforms fall in several categories. See generally *Developments in the Law—The Civil Jury*, 110 HARV. L. REV. 1408, 1527-36 (1997) (describing state reform efforts to regulate punitive damages). Some states have enacted caps, see, e.g., *id.* at 1533 & n.158; other states are considering caps, see, e.g., *id.* Some caps limit the entire dollar amount, see, e.g., *id.* at 1533 & n.159; others limit punitive damage awards to a multiple of compensatory damages, see, e.g., *id.* at 1533 & n.160; others cap possible punitive amounts at three times the amount of compensatory damages, see, e.g., FLA. STAT. ch. 768.73(2) (1995). Some states require part or all of the punitive award to go to state agencies or the state treasury. See, e.g., KAN. STAT. ANN. § 60-3402(e) (1994). Connecticut and Kansas have required the judge, not the jury, to determine punitive damage awards. See CONN. GEN. STAT. § 52-240b (1997); KAN. STAT. ANN. § 60-3701(a) (Supp. 1996). Some states have required a bifurcated trial designed to reduce confusion and to make clear which factors are relevant to which proceeding. See, e.g., *Developments in the Law—The Civil Jury*, *supra*, at 1529-31. Over half the states have increased the standard of proof from preponderance of the evidence to clear-and-convincing evidence or more. See, e.g., *id.* at 1531-33.

111. They are crude because no one believes that a simple ratio or flat cap makes much sense except as an easily administrable effort to ensure against the most outrageous awards.



A total of 899 respondents were selected at random from the voter registration rolls of Travis County, Texas,<sup>112</sup> and paid \$35 to each participant in the study. We created a set of ten vignettes of personal injury cases in which a plaintiff (always an individual customer) sued a firm for compensatory and punitive damages. For each vignette, we constructed versions that differed in the size of the defendant firm (medium or large). For four of the vignettes, there were also versions that differed in the harm that the plaintiff was said to have suffered, but not in the description of the defendant's actions. In total, there were twenty-eight different variations of the ten vignettes. Each respondent evaluated ten cases, composed of one variation of each of the ten basic vignettes. The respondents were told to assume in all cases that compensatory damages had been awarded in the amount of \$200,000, and that punitive damages were to be considered. Three subsamples were asked to answer different questions about each scenario: how outrageous the defendant's behavior was (on a scale of 0 to 6); how much the defendant should be punished (on a scale of 0 to 6); and how much the defendant should be required to pay in punitive damages. Each respondent first dealt with one case without seeing the others (for these evaluations, we will say that the "no-comparison" condition applied), then received a booklet with nine new cases. A further description of the study, an excerpt from the instructions, a recitation of the scenarios, and other information are contained in the appendices below.

#### A. *Shared Outrage*

We propose a descriptive theory of the psychology of punitive awards, called the outrage model.<sup>113</sup> The essential claim is that the moral transgressions of others evoke an attitude of outrage, which combines an emotional evaluation and a response tendency. The rules that govern outrage present an important problem that we do not address in this Article. We assume that outrage is governed largely by social norms. Judged by reference to these norms, a particular person's expressions of indignation may be deemed too intense for its cause or not intense enough. Social as well as legal norms also regulate the mapping from transgressions to punishment.

An attitude is a mental state and is not directly observable. The various aspects of an attitude can, however, be "mapped" onto diverse responses, which might include facial expressions, verbal statements of opinion, gestures—even physical assault. Response "modes" might include a judgment about outrageousness on a numerical scale. Under the outrage model, punitive

---

112. Travis County was selected for convenience. Travis County is a favorite of national market researchers precisely because of its diversity. It is the most liberal county in a conservative state and has considerable racial and income diversity, as reflected in our sample. It is conceivable but unlikely that the shared judgments that we find would not generalize to, for example, the United States in general.

113. For further discussion, see Kahneman et al., *supra* note 1 (manuscript at 49-53).

damages are considered an expression of an angry or indignant attitude toward a transgressor. The evaluative aspect of the attitude is labeled outrage; the response tendency is labeled punitive intent. Outrage is basic, and punitive intent is measured by outrage and additional factors, such as harm. As we will see, the verbal indication of the desired severity of punishment (punitive intent) is affected both by the outrage that an action evokes and by the severity of its consequences ("harm"). This retributive aspect of punishment is incorporated in many aspects of the law, such as the large discrepancy between sentences for murder and for attempted murder. The relationship between victim and juror was manipulated in an experiment by Reid Hastie, David Schkade, and John Payne.<sup>114</sup> Under experimental conditions, larger awards were made when the defendant was from a remote location. We speculate that this factor affects punitive intent: The retributive urge is stronger when the victim belongs to one's group than when the victim is a stranger.<sup>115</sup>

In some situations the expression of an attitude is restricted to a particular scale of responses. For example, in the situation with which we are concerned here, the responses of juries are restricted to a scale of dollars. Dollar amounts of punitive awards (like the length of prison sentences in criminal cases) are only one of a number of possible scales on which outrage might be expressed. By adopting a particular response "mode," the legal system, in a sense, constructs the relevant values. A bounded numerical scale would be another obvious possibility. We propose that some factors affect the mapping of punitive intent onto the dollar scale, but that these factors do not affect punitive intent on a bounded numerical scale. For example, the size of the defendant firm is an important factor in translating punitive intent into dollars; a judgment that appears severe when the defendant is a small firm may appear grossly inadequate when the defendant is a giant. Thus firm size will affect dollar awards even if it does not affect punitive intent as measured on a bounded numerical scale.

In summary, the outrage model assumes an internal state of outrage, which can be mapped onto different response scales. These scales vary not only in their complexity, but also in the precision and reliability of the measurement that they support: Some scales are "noisier" than others. As we will shortly see, the dollar scale is an extremely noisy expression of punitive intent.

---

114. See 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 21-22* (1998) (unpublished manuscript, on file with the *Yale Law Journal*).

115. The amount demanded by plaintiff also affected the size of the awards. This observation is most likely to be an anchoring effect, which influences the dollar award directly, independently of punitive intent. The anchoring effect may be quite important in light of the unfamiliarity of the dollar scale as a scale of punishment, a point that we will discuss in some detail. See *infra* Subsection IV.A.2.

## B. *Hypotheses*

Our two central hypotheses were that jurors would exhibit shared outrage, but that dollar awards would be erratic. The basis for the shared outrage hypothesis is the belief that a bedrock of broadly shared social norms governs the outrage evoked by different scenarios of tortious behavior. With respect to outrage and punitive intent, we thus predicted that randomly selected juries are likely to be similar to one another. In addition, we predicted that rankings of different scenarios would be generally similar for different demographic groups, at least in the context of the personal injury cases given here. The erratic dollar amounts hypothesis was that, in contrast to outrage and punitive intent, which are measured on bounded scales, punitive awards denominated in dollars are susceptible to large individual differences, which could be a significant cause of the unpredictability of jury determinations. We tested this prediction by comparing the extent of variations in judgments about outrageousness and appropriate punishment with the extent of variations in judgments about appropriate dollar awards.

We also examined three other hypotheses. First, we considered the “harm effect.” We hypothesized that punitive intent—as measured on a bounded numerical scale—is determined by the outrageousness of the defendant’s behavior and by other factors, prominently including the harm suffered by the defendant. The prediction that harm affects punitive intent but not outrage was tested by presenting alternative versions of some scenarios, in which the severity of the harm suffered by the plaintiff varied. Second, we tested whether the defendant firm’s size was relevant. We hypothesized that damage awards are determined by punitive intent, and by other ascertainable factors, prominently including the size of the defendant firm. We tested this prediction by presenting each scenario in two versions, in which the size of the defendant varied. Third, we considered the effect of having comparison cases. In a “no-comparison” condition, we hypothesized that there is a cautious tendency to place cases toward the middle of a bounded numerical scale, and hence people will not make appropriate distinctions among cases. We tested this hypothesis by giving respondents cases in isolation and in the context of other cases.

## C. *Results*

### 1. *Shared Outrage*

Our first question was whether the degree of outrage is consistent across individuals and across possible juries. A simple way to answer this question is to examine whether rankings of different scenarios are generally similar for

different demographic groups. We therefore computed the means of the three responses (outrage, punitive intent, and dollar awards) separately for groups of respondents defined by demographic variables (men, women, whites, Hispanics, blacks, different levels of income and education). To measure the level of agreement across disjoint categories (e.g., men and women), we computed the correlation between their average responses<sup>116</sup> over the set of twenty-eight cases.

The correlations were remarkably high for judgments both of outrage and of punitive intent. In particular, there was essentially perfect agreement among groups in the ranking of cases by punitive intent: the median correlation was .99. Men and women, Hispanics, blacks, and whites; and respondents at very different levels of income and education produced almost identical orderings of the twenty-eight scenarios used in the study.<sup>117</sup> Judgments of intent to punish in these personal injury scenarios evidently rest on a bedrock of moral intuitions that are broadly shared in society.<sup>118</sup> We also looked for differences among groups in the average severity of their judgments (in absolute terms, as opposed to their rankings), and here too found no significant differences (with the one exception of gender, to be discussed below<sup>119</sup>). Table 1 summarizes our findings.

---

116. We used means for outrage and punishment and medians for dollar awards. The means and medians of outrage and punishment ratings are very close together because their distributions are essentially symmetric, and thus the same answer emerges either way. Dollar awards have a severely right-skewed distribution, and means are thus heavily influenced by a few large outliers while medians are much more stable. For example, median dollar awards for a given scenario range from \$0 to \$2 million, while mean dollar awards range from \$89,000 to \$22.1 million. We performed all key analyses using both means and medians of dollar awards, and in all cases means make dollar awards perform even more unpredictably than with the less erratic medians. We used medians to be conservative about our claims of excessive variability in dollar awards. Moreover, our normative argument about community sentiment is based on a hypothetical median voter, providing another reason for preferring median measures over means.

117. Of course, considerable variability among individuals remains within a given demographic category, even though the aggregate responses of one category are very similar to the aggregate responses of another category. Our analysis simply shows that identifiable groups have similar distributions, not that all individuals are alike.

118. There is a considerable amount of individual variability even in outrage and punishment ratings, as there always is in any evaluative judgment task. Only 49% of the variance in individual punishment ratings, 29% of outrage ratings, and 6% of dollar awards are explained by differences between cases. See Kahneman et al., *supra* note 1 (manuscript at 38 tbl.6). From one point of view, 49% is evidence of a lack of consensus. In another view, however, it is a remarkably high level of agreement across a wide diversity of people in how they rated (not just ranked) 10 different cases. What is impressive is that even a small aggregate like a jury, which is composed of otherwise diverse individuals, so consistently converges on a similar result. When we speak of shared outrage, we mean not that social norms are so powerful that all individuals agree, but rather that they are shared widely enough that even a modest search (e.g., asking a few randomly selected people) quickly reveals the center of society's judgment.

119. See *infra* text accompanying notes 122-126.

TABLE 1. CORRELATIONS AMONG DEMOGRAPHIC GROUPS OF INTENDED SEVERITY OF PUNISHMENT<sup>120</sup>

GENDER

	Men
Women	.974

ETHNICITY

	Black	White
White	.975	
Hispanic	.963	.988

HOUSEHOLD INCOME

	< \$30,000	\$30,000-\$50,000
\$30,000-\$50,000	.991	
> \$50,000	.986	.986

AGE

	< 30	30-39	40-49
30-39	.994		
40-49	.992	.994	
> 50	.991	.993	.987

This striking finding may not generalize to all domains of the law.<sup>121</sup> We might expect to find larger differences between communities and social categories in other domains, perhaps including attitudes toward civil rights violations and environmental harms. It is at least possible that, for example,

120. Entries are correlations between mean responses to scenarios by respondents in the indicated demographic categories.

121. In addition, the competing narratives that real world juries hear might create more heterogeneity than is revealed by the responses to our fairly stark case descriptions. In some cases, it is imaginable that members of different groups would be especially alert to some facet of the narrative of one or another side. Experiments involving mock juries might be designed to test for this possibility. (We are now engaged in experiments of this kind.)

blacks would rate civil rights violations more severely than whites do. We expect, however, that correlations among rankings are likely to remain the same across different groups, so that different demographic groups would agree on which defendants have behaved least and most egregiously. Here there is a great deal of room for further empirical work.

As noted, the only statistically significant difference in average ratings was between women and men. While women and men ranked the scenarios identically (as indicated by the extremely high correlations), men were somewhat more lenient and women were somewhat more severe: Women rated the defendant's behavior as more outrageous,<sup>122</sup> expressed more punitive intent,<sup>123</sup> and set higher log dollar awards.<sup>124</sup> (It would be equally accurate to say that men rated behavior as less outrageous, expressed less punitive intent, and set lower awards.) There was also an interaction between scenarios and gender, in which women assigned even higher ratings of outrage and punishment (but not higher awards) to cases in which the plaintiff was female—or, alternatively put, in which men assigned lower ratings in such cases.<sup>125</sup>

These findings should not be overemphasized; the differences in ratings, while statistically significant, were relatively small. These findings do, however, bear on legal and social disputes about jury composition, providing some empirical evidence that women do reach different conclusions from those reached by men, and in particular that they seek more severe punishment of civil defendants. The finding that women tend to favor stiffer punishments—and that men tend to favor more lenient ones—is in one sense counter to folk wisdom, which sees women as particularly lenient. But our finding is in line, broadly speaking, with other research suggesting that women and men view social risks differently and, in particular, that women tend to view such risks as more serious than do men.<sup>126</sup>

## 2. *Unpredictable Dollar Awards*

With respect to dollar awards, our central hypothesis was that such awards are erratic because of individual differences in the mapping of punitive intent onto the dollar scale. To test this hypothesis we produced simulated juries by randomly sampling, with replacement, groups of twelve responses to each case for each response scale. In this manner we

---

122. The mean difference between women and men was .52 scale units, with a statistical significance of  $p < .001$ .

123. The mean difference was .37 scale units ( $p < .001$ ).

124. This finding had statistical significance of  $p < .01$ .

125. This finding had a statistical significance of  $p < .05$ .

126. See Paul Slovic, *Trust, Emotion, Sex, Politics, and Science: Surveying the Risk Assessment Battlefield*, 1997 U. CHI. LEGAL F. 59, 71, 73-74.

constructed a large number of three types of simulated juries: "outrage juries," "punishment juries," and "dollar juries." Of course there is a large question about how a set of individual judgments will produce a jury verdict. No doubt group dynamics can push deliberations in unexpected directions, sometimes toward the most extreme member of the group. As a statistical matter, however, the experimental literature on the relationship between prior individual judgments and the outcomes of group deliberation suggests that the median judgment is a good predictor.<sup>127</sup> Indeed, using the median judgment may even understate our ultimate conclusion about the variability of dollar awards.<sup>128</sup>

We therefore used the median judgment of each simulated jury as an estimate of what the judgment of that jury would have been. Without losing sight of the limitations of our estimation procedure, we apply the label "jury judgment" to these estimates for simplicity of exposition. Table 2 summarizes the simulated jury judgments for punishment and dollar awards.

Jury judgments can be considered shared and therefore predictable, in our use of that term, if there is high agreement among juries randomly selected from the population. In order to find a source of erratic judgments, we attempted to compare the predictability of the judgments made by simulated dollar juries, outrage juries, and punishment juries. First, we imagined that all of our case scenarios were tried on the same day by independent juries, analogous to how jury judgments for different cases are produced in practice. We then asked the question, "If these same cases were tried again independently, how likely would we be to get the same ratings and rankings as in this first set of trials?"

---

127. See James H. Davis, *Group Decision Making and Quantitative Judgments: A Consensus Model*, in UNDERSTANDING GROUP BEHAVIOR 35, 47 (Erich Witte & James H. Davis eds., 1996). Clearly, the appropriateness of this measure may depend on the task structure of the group (e.g., whether a unanimous decision is required). In our study, replacing the median of jurors' individual judgments as the group decision with the mean had little effect on the results for outrage and punishment, but made dollar awards even more unpredictable.

128. See *id.* at 48-49; Shari Diamond & Jonathan Casper, *Blindfolding the Jury to Verdict Consequences*, 26 L. & SOC'Y REV. 513, 553 (1992); Martin F. Kaplan & Charles E. Miller, *Group Decision Making and Normative Versus Informational Influence*, 53 J. PERSONALITY & SOC. PSYCHOL. 306, 309 (1987). Note, however, the existence of an "amplification of bias," by which a group process, involving a set of individuals biased in one direction or another, may push awards in extreme directions, in fact more extreme than that of any individual before deliberation begins. Cf. Norbert L. Kerr et al., *Bias in Judgment: Comparing Individuals and Groups*, 103 PSYCHOL. REV. 687 (1996) (finding an amplification of bias in settings other than punitive damage determinations). We are indebted to Robert MacCoun for suggesting this possibility. The possibility of extremes resulting from group deliberation would fortify our conclusion by showing even greater variance.

TABLE 2. SYNTHETIC JURY RESPONSE DISTRIBUTIONS BY SCENARIO

SCENARIO	FIRM SIZE	HARM LEVEL	LOWER 95% CONFIDENCE BOUND	MEDIAN	UPPER 95% CONFIDENCE BOUND	MEAN JURY PUNISHMENT	PREDICTION ERROR RATIO (\$/PUNISHMENT)
Joan	Large	High	\$500,000	\$2,000,000	\$15,000,000	5.14	3.36
Joan	Medium	High	200,000	900,000	3,000,000	5.03	2.27
Thomas	Medium	————	200,000	500,000	1,575,000	5.02	1.69
Martin	Medium	High	350,000	1,000,000	4,000,000	4.98	4.01
Thomas	Large	————	200,000	560,000	2,750,000	4.95	0.50
Joan	Large	Low	175,000	1,000,000	12,500,000	4.93	13.57
Martin	Large	High	350,000	1,900,000	10,000,000	4.92	2.40
Frank	Medium	————	230,000	760,000	2,100,000	4.86	1.67
Frank	Large	————	225,000	1,000,000	4,000,000	4.82	2.62
Mary	Large	————	290,000	1,000,000	4,000,000	4.79	1.49
Joan	Medium	Low	150,000	750,000	5,500,000	4.71	9.90
Mary	Medium	————	250,000	710,000	2,100,000	4.70	2.51
Martin	Large	Low	350,000	1,000,000	5,000,000	4.47	3.63
Martin	Medium	Low	200,000	675,000	2,250,000	4.16	2.53
Susan	Large	————	100,000	300,000	1,000,000	3.27	1.78
Susan	Medium	————	50,000	225,000	800,000	3.03	0.93
Janet	Medium	High	100,000	200,000	690,000	2.79	1.37
Carl	Medium	————	15,000	155,000	375,000	2.78	1.59
Carl	Large	————	50,000	200,000	750,000	2.64	1.59
Janet	Medium	Low	0	150,000	650,000	2.49	2.00
Janet	Large	High	0	287,500	1,500,000	2.39	4.41
Janet	Large	Low	12,500	200,000	1,000,000	2.38	1.30
Jack	Large	High	0	0	350,000	1.24	2.10
Jack	Medium	High	0	45,000	225,000	1.07	1.30
Jack	Medium	Low	0	0	112,500	1.03	0.89
Jack	Large	Low	0	2550	500,000	0.95	2.91
Sarah	Large	————	0	0	1000	0.51	1.12
Sarah	Medium	————	0	0	13,000	0.46	∞
MEDIAN							2.18

To answer this question we conducted an analysis requiring four steps. In Step One, we created a randomly selected jury for each of the twenty-eight cases, and computed the median judgment for each. In Step Two, we imagined that a time machine allowed us to replay each case again independently of the first trial, including the random selection of a new jury. We therefore created a second set of twenty-eight randomly selected juries and corresponding median judgments. The correlation between the second set and the first set of jury judgments is a measure of how erratic or consistent juries are. In Step



Three, we performed Step One sixty times for each of the three response modes, and repeated Step Two to get a more reliable indication of the typical correlation between juries. This produced 180 columns of data, each of which contained one set of twenty-eight jury judgments. Finally, we computed the correlations between every pair of sets of simulated jury judgments (i.e., correlations between the columns). This computation was performed both within responses (e.g., the correlation between any two sets of twenty-eight outrage ratings) and across responses (e.g., the correlation between a set of twenty-eight outrage ratings and a set of twenty-eight punishment ratings). The data shown in Table 3 are medians of the correlations<sup>129</sup> obtained for each response mode or response mode combination.

TABLE 3. MEDIAN CORRELATIONS BETWEEN SETS OF SIMULATED JURIES

	OUTRAGE	PUNISHMENT	\$ AWARDS
OUTRAGE	.87		
PUNISHMENT	.86	.89	
\$ AWARDS	.47	.51	.42
OVERALL MEDIAN AWARD	.71	.77	.69

As we had hypothesized, the individual differences in dollar awards produce severe unpredictability and highly erratic outcomes, even in the medians of twelve judgments (the results would be even more extreme with smaller samples such as six-person juries). While there is strong agreement between independent sets of outrage or punishment juries (median  $r = .87$  and  $.89$ , respectively), agreement between independent sets of dollar juries is quite weak (median  $r = .42$ ).<sup>130</sup> The variability of individual dollar judgments is so large that even the medians of the judgments of twelve-member juries are quite unstable. The problem could be reduced, of course, by taking larger samples. For example, and significantly, we found that the correlation between sets of dollar juries rises to  $.80$  when the size of the juries is increased to thirty (and correlations for thirty-person outrage and punishment juries rise even higher, to  $.95$  and  $.97$ , respectively).

129. There is nothing magic about medians here except that the mean of a set of correlations is not very meaningful from a technical viewpoint, whereas the median is readily interpretable. Here it makes little difference: Means are about  $.01$  lower than medians.

130. The " $r$ "-value refers to correlations. Thus the data shown in Table 3 are the medians of the 1770 correlations obtained within each response mode or of the 3600 correlations obtained between two response modes. There is no direct interpretation of a raw correlation (unless it is 0 or 1), but squaring it produces the percentage of variance in one variable explained by the other. Thus only 18% of the variance in one set of dollar juries ( $0.42$  squared) is predictable from another set (and the remaining 82% is unexplained).

### 3. *The Effects of Context, Harm, and Firm Size*

Unlike real jurors, who are exposed to a single case for a long time, the participants in our study responded to a total of ten personal injury cases in quick succession and had an opportunity to compare most of these scenarios to one another. To examine the effect of this unusual procedure, every participant first encountered one of the first six scenarios, which was presented in a separate envelope and was evaluated in isolation from the others. Our basic finding was that in this no-comparison condition, there was a cautious tendency toward diminished differentiation in judgments about different cases.<sup>131</sup> The availability of comparison cases apparently makes a serious case appear more serious than it would on its own and makes a mild case appear milder. Thus the most consistent effect of context is to increase the *range* of the judgments across different scenarios. We conclude that the availability of a context of similar cases improves people's ability to discriminate among these cases, but does not affect the basic moral intuitions that the cases evoke.

We also tested whether the consequences of a defendant's actions were important to lay intuitions about proper punishment, by constructing alternative versions of four additional vignettes that differed in the harm that the plaintiff had suffered.<sup>132</sup> As predicted, we found that the harm that occurred did not affect the degree of outrage evoked by the defendant's behavior.<sup>133</sup> In contrast, varying the harm had a small but statistically significant effect on punishment ratings, where defendants who had done more harm to the plaintiff were judged to deserve greater punishment. This harm effect carried through

---

131. To test whether the availability of a context altered intuitions, we computed rank-correlations ( $n = 12$ ) over the scenarios, between the judgments in the isolation condition and in the context condition. The correlations are high: .88 and .90, respectively, for the means of outrage and punishment ratings, and .89 for the medians of dollar awards. The availability of a context of comparison slightly affected the distribution of judgments by increasing the range of the judgments of different scenarios. The variances of mean judgments of outrage and punishment over the 12 scenarios are larger in the context condition for punishment ( $F(1,10) = 6.97, p < .05$ ) and are marginally significant for outrage ( $F = 4.07, p < .10$ ) and median dollar awards ( $F = 4.64, p < .10$ ).

We are now engaged in a study designed to test for consistency in jury judgments and in particular for the possibility of judgment reversals, as where case *A* produces a higher award than case *B* in isolation, but case *B* a higher award than case *A* when the two are compared. Cf. Richard Thaler & Amos Tversky, *Preference Reversals*, in RICHARD THALER, *THE WINNER'S CURSE* 79 (1992) (discussing related phenomena).

132. Note that the difference was measured qualitatively rather than in dollar terms; in all of the cases, the jury had awarded \$200,000 in compensatory damages, but in some of them the description of the injury suggested variations in the way of qualitative loss, as for example in the cases involving a burned child.

133. An action can be judged more or less outrageous without reference to its consequences; certainly it is possible to think that the outrageousness of an action does not depend on what actually happened. Consequences, however, are important to punishment in law. There is a great deal of discussion about whether attempted crimes should be punished less severely than well-executed crimes if both show the same state of mind and if an unsuccessful attempt failed because of some accident. See, e.g., Steven Shavell, *Deterrence and the Punishment of Attempts*, 19 J. LEGAL STUD. 435 (1990).

to dollar awards. Low harm produced an average award of \$727,599, and high harm, an average award of \$1,171,251.

Finally, we hypothesized that firm size would affect neither outrage nor punitive intent, but that the same degree of punitive intent would be translated into a larger amount of damages when the firm is larger than when it is smaller. As expected, we found no statistically significant effects of firm size on either outrage or punishment judgments. But large firms were punished with much larger dollar awards (an average of \$1,009,994) than medium-sized firms (\$526,398).<sup>134</sup> This is substantial evidence that equivalent outrage and punitive intent will produce significantly higher dollar awards against wealthy defendants.<sup>135</sup>

#### D. Discussion

How do these findings bear on the appropriate role of juries in setting punitive damage awards? As we have seen, a conventional understanding of such awards sees the jury as a sample from the community whose function is to provide an estimate of community sentiment. If jury judgments are erratic, this function is badly compromised, *for any particular jury's judgment may not reflect community sentiment at all*. The bottom row of Table 3 presents the median correlations between sets of simulated jury judgments for the twenty-eight scenarios and the corresponding estimates of community sentiment, for which we used the overall median of dollar awards for each scenario. It is obvious that the judgments of dollar juries provide a poor estimate of overall community sentiment. Indeed, the unreliability of dollar juries is so pronounced that the dollar awards that would be set by the larger community are predicted more accurately by punishment juries. This is a counterintuitive finding. It leads directly to a possible recommendation, which we explore below: Juries instructed to state their punitive intent could be used, in conjunction with a

---

134. Analysis of variance yields a significant effect of firm size on log awards ( $F = 4.8, p = .03$ ) and also on raw dollar awards ( $F = 6.1, p = .01$ ).

135. Within the academic community, opinion is sharply divided on the question of whether the amount of punitive awards should depend on the size of the defendant firm. See, e.g., Abraham & Jeffries, *supra* note 45, at 415-19 (arguing that defendant's wealth is irrelevant); Dan B. Dobbs, *Ending Punishment in "Punitive" Damages: Deterrence-Measured Remedies*, 40 ALA. L. REV. 831, 871-72 (1989) (arguing that wealth is "relevant only when the defendant's conduct was noneconomic in nature and motive"); Galanter & Luban, *supra* note 8, at 1431-32 (arguing that "punitive damages should be scaled to the defendant's wealth"); Polinsky & Shavell, *supra* note 7, at 910-14 (arguing that defendant's wealth should not matter). Lay intuitions, in contrast, are quite clear. A psychological analysis suggests that people think in terms of retribution rather than deterrence and that the intention to punish is an intention to inflict pain; this means that the size of the defendant matters a good deal.

We do not deny that there is a plausible account of deterrence that would make firm size pertinent. That account might stress the organizational structure of the firm and suggest that high-level managers will not alter policies unless an award is sufficient to "get their attention." Polinsky and Shavell suggest that this view is implausible, but that question cannot be resolved a priori; it is an empirical issue. See Polinsky & Shavell, *supra* note 7, at 912. MacCoun, by comparison, finds effects from corporate identity and commercial activity, but not from wealth per se. See MacCoun, *supra* note 46, at 133-43.

preset conversion function, to generate punitive awards that would more accurately represent community sentiment, thus reducing much of the unpredictability of awards.<sup>136</sup>

1. *The Underlying Problem: "Scaling Without a Modulus"*

A key to our analysis is a distinction that psychologists draw between two types of scales. *Category scales* are bounded and anchored in verbal descriptions at both ends. Scales of this type are often used in public opinion surveys and were used here to measure outrage and punitive intent. *Magnitude scales*, by contrast, are unbounded and are defined by a meaningful zero point. These scales are often used in the psychological laboratory, for example to scale the brightness of lights or the loudness of sounds. Magnitude scales have occasionally been used to measure the intensity of responses to socially relevant stimuli, such as the severity of crimes and the severity of punishments.<sup>137</sup> The dollar scale of punitive awards is obviously not a category scale; it satisfies the defining characteristics of a magnitude scale, for the zero point is meaningful and the scale is unbounded.

Although the relations between the two types of scales have been the subject of much scholarly attention and controversy,<sup>138</sup> some characteristic differences between them are well-established. First, the distributions of judgments on magnitude scales are generally positively skewed, with a long right tail; this is a consequence of the fact that the scale is bounded by zero at the low end. Second, judgments on magnitude scales are often highly variable. Third, the standard deviations of individual judgments of different objects (a measure of variability) is often roughly proportional to the mean judgments of these objects.

The common practice in laboratory uses of magnitude scaling is to define a *modulus*: Respondents are instructed to assign a particular rating to a "standard" stimulus, defined as the modulus, and to assign ratings to other stimuli in relation to that modulus. Thus, for example, a modulus of "5" might be assigned to a noise of a certain volume, and other noises might be assessed in volume by comparison with the modulus. An experiment can, however, be conducted without specifying a modulus. In this situation of magnitude scaling without a modulus, different respondents spontaneously adopt different moduli, but their responses generally preserve the same ratios even when the moduli differ. For example, one observer may assign a judgment of 200 to a stimulus

---

136. See *infra* notes 158-161 and accompanying text.

137. See STANLEY S. STEVENS, *PSYCHOPHYSICS* 252-58 (1975).

138. See, e.g., *id.* at 26-31; R. Duncan Luce & Carol L. Krumhansl, *Measurement, Scaling, and Psychophysics*, in 1 STEVENS' HANDBOOK OF EXPERIMENTAL PSYCHOLOGY 3 (Richard C. Atkinson et al. eds., 2d ed. 1988); Lawrence E. Marks, *Magnitude Estimation and Sensory Matching*, 43 PERCEPTION & PSYCHOPHYSICS 511 (1988).

that another observer rates as 10. If the first observer now assigns a rating of 500 to a new stimulus, we may expect the second to assign to that stimulus a value of roughly 25.

Here is the central point: Magnitude scaling without a modulus produces extremely large variability in judgments of any particular stimulus, because of arbitrary individual differences in moduli. The assignment of punitive damages satisfies the technical definition of magnitude scaling without a modulus.<sup>139</sup> This reasoning led us to adopt the central hypothesis of the present study described and established above.<sup>140</sup> In an important sense, the current system for awarding punitive damages is analogous to what Max Weber described as “khadi justice,”<sup>141</sup> a highly particularistic legal regime in which one outcome need not predict any other.

## 2. A Simple Way To Improve Predictability

A conventional view about the role of juries in setting punitive awards is that the jury is a sample from the community, whose function is to provide an estimate of community sentiment. In the context of our experiment, community sentiment about the punitive damages for a scenario was defined as the median of the damages set by all the respondents who judged it. This sentiment represents population-wide judgments about appropriate dollar awards. The findings summarized in Table 3 suggest a straightforward procedure for improving the accuracy with which this community sentiment can be estimated from the judgments of a sample of twelve citizens: judgments of punitive intent and a conversion function based on the results of a large sample, one that can be taken to reflect a population-wide judgment. To measure the accuracy of a simulated jury as an estimate of the population median, we

---

139. In the legal context, some moduli might be insufficiently informative even if provided. If, for example, \$0 means entirely acceptable, and \$200 means objectionable (as, for example, in the case of a reckless act causing minimal harm), the jury would probably continue to be at a loss for most punitive damage cases. A modulus would have to provide a standard around which judgments could be manageably organized in the likely comparison set. In principle, the dollar scale is more familiar and therefore more understandable to jurors. The range of dollar amounts with which most citizens have any experience (generally in the thousands or tens of thousands), however, is dwarfed by the magnitudes that are considered in scenarios such as ours (hundreds of thousands or millions and up), and thus for all practical purposes they are dealing in an imaginary range of dollar amounts.

140. Similar evidence emerges from Michael J. Saks et al., *Reducing Variability in Civil Jury Awards*, LAW & HUM. BEHAV., Spring 1997, at 243. Saks et al. show that under experimental conditions, variability can be reduced in the context of pain-and-suffering awards by providing people with information about average dollar awards for the type of injury at issue, intervals (i.e., where 80% of awards for similar injuries fell), and comparable examples (awards for four similar cases). *See id.* at 249. Without using the concepts, Saks et al. in effect supply a modulus in each of these conditions, and the result is to decrease variability dramatically. *See id.* at 250-51. See also the proposal for a form of scaling, through baseline appraisals based on statistical profiles, in Glen O. Robinson & Kenneth S. Abraham, *Collective Justice in Tort Law*, 78 VA. L. REV. 1481, 1490 (1992).

141. 3 MAX WEBER, *ECONOMY AND SOCIETY: AN OUTLINE OF INTERPRETIVE SOCIOLOGY* 976-78 (Guenther Roth & Claus Wittich eds., Bidminster Press 1968).

computed the discrepancy between the dollar award set by a jury (for punishment juries, we used the dollar value from the conversion function) and the overall median dollar award for that case.<sup>142</sup> From these discrepancies we can compute the root mean squared error (RMSE), which is a conventional measure of accuracy of estimation and is analogous to the standard error of the estimate in a regression model.

This analysis provides two values of RMSE for each of the twenty-eight cases in our study, one from dollar juries and one from punishment juries. In twenty-five of the twenty-eight cases, RMSE is smaller (indicating higher accuracy) for estimates derived from punishment juries. To assess the magnitude of the effect, we computed the ratio of the two values of RMSE for each scenario (listed in the last column of Table 2). The median ratio was 2.18, which means that for the median case, using dollar juries leads to over twice as much prediction error than using punishment juries and a conversion function. For example, for the case of Joan with a medium firm size and high harm, the ratio is 2.27, the median award is \$1,000,000, and the estimates of dollar juries have an average error from this value of \$913,481 compared to \$402,414 for estimates based on punishment juries.

In simple language, this means that unpredictability could be greatly reduced, and population-wide judgments about dollar awards would be far more likely to be found, if the legal system dispensed with dollar juries and relied instead on punishment juries and a conversion function. Indeed, as can be seen in Table 2, the median probably understates the decisive overall advantage of using predictions based on punishment ratings, since for some individual cases the reduction in error is extremely large.

One note of caution is in order here. The fact that punitive damages share the known deficiencies of magnitude scaling is likely to be a significant cause of unpredictable punitive awards—but it is not the only one. Other factors include regional differences, plaintiffs' demands, anchors' of various sorts, differences in social norms over both time and space, the quality of the lawyers on both sides, and doubtless others. We take up some of these points in the next part.

---

142. We first estimated the conversion function separately for medium firms and for large firms, since there was a significant effect of firm size on dollar awards. Following Stevens, *see* STEVENS, *supra* note 137, at 100-01, we estimated power functions, which expressed our measure of overall community sentiment for a given scenario (the median of individual dollar awards) as a function of the mean judgment of punishment juries. We then generated two sets of simulated jury judgments for each case. The first set consisted of the median judgments of 100 randomly sampled dollar juries. The second set was obtained by taking the median judgments of 100 randomly selected punishment juries, and by transforming this value to dollars separately for each jury by the appropriate conversion function (depending on whether the defendant firm was large or medium-sized).

#### IV. POLICY REFORMS: COMMUNITY JUDGMENTS WITH MEANING AND WITHOUT NOISE?

##### A. *Preliminary Observations on Limitations and Reform*

###### 1. *Experiments and the Real World*

What are the implications of our findings for punitive damage reform? Before answering that question, it is necessary to emphasize that these findings do not replicate the real world of punitive damage awards. Ours was an experimental study, and our juries consisted of individuals who were given brief narrative descriptions of cases. They were not presented with full accounts, much less with adversarial arguments on both sides. These arguments could introduce additional variance; they could tend to reduce disparities. The fact that lawyers on both sides can typically exclude certain jurors may reduce the degree of variance in real world awards, at least if lawyers can anticipate which people will have outlier moduli. Moreover, our "juries" did not deliberate.<sup>143</sup> As a statistical matter, the median vote is not a bad prediction, but it is certainly a crude one, at least for individual cases.

###### 2. *Anchors and Their Effects*

Our study referred to compensatory damages and to firm size, but it did not contain two usual "anchors": the plaintiff's demand and the jury's own prior determination of compensatory damages. Such anchors are likely to matter a great deal in producing actual awards.<sup>144</sup> Other anchors may well emerge during the lawyers' advocacy. Notably, there is evidence of a correlation between anchoring and confidence; that is, when people are not confident of their judgments, they are more susceptible to anchoring effects.<sup>145</sup> We suspect that people are not confident of the dollar amounts they award,<sup>146</sup> and that this lack of confidence makes the use of anchors more likely. It is natural to think that the process of magnitude scaling without a modulus will encourage people to seize on whatever anchors are available,

---

143. We are now embarking on a follow-up study that involves groups of mock juries and that thus attempts to be more precise about the effects of deliberation.

144. For experimental evidence that the plaintiff's demand has considerable importance, see Gretchen B. Chapman & Brian H. Bornstein, *The More You Ask for, the More You Get: Personal Jury Verdicts*, 10 APPLIED COGNITIVE PSYCHOL. 519, 537 (1996). For evidence about anchors in awarding damages for pain and suffering, see Saks et al., *supra* note 140, at 253-55. For real world evidence about the anchoring effects of compensatory awards, see Eisenberg et al., *supra* note 21, at 637.

145. See Karen E. Jacowitz & Daniel Kahneman, *Measures of Anchoring in Estimation Tasks*, 21 PERSONALITY & SOC. PSYCHOL. BULL. 1161, 1164-65 (1995).

146. Those who participated in our study said that the most difficult part of their task was the assignment of dollar amounts.

regardless of whether they are sensible from the point of view of the goals of the legal system.

How does the existence of real world anchors affect our findings? The answer is that anchors may or may not increase predictability. If it is hard to know in advance what will be used as the anchor, predictability will be absent; if everyone knows what the anchor is likely to be, there will be fewer unpredictable outcomes. But if this is so, predictability comes with a cost of its own: It introduces an additional layer of arbitrariness, if, as is likely, the anchor is itself arbitrary on normative grounds. There is no reason to think that the plaintiff's demand should carry a great deal of weight in determining the proper punitive award. And if the compensatory award anchors the punitive award, there is a kind of arbitrariness to the extent that anchor is arbitrary, as deterrence theory suggests that it is. We suspect that juries lacking evident anchors will suffer from the problems we have described, whereas juries resorting to anchors will produce arbitrariness of a different sort.

### 3. *The Rule of Law and How To Obtain (and How Not To Obtain) Its Virtues*

Many concerns about punitive damage awards focus on their apparent arbitrariness, and many proposed remedies attempt to promote rule-of-law values through, for example, more careful and more specific judicial instructions. In *BMW*, Justice Breyer spoke in some detail of the failure of Alabama law sufficiently to discipline jury discretion with clear criteria about the grounds of awards.<sup>147</sup> Thus a likely response to complaints about arbitrary awards is to increase the specificity of instructions to juries.

Our study strongly supports Justice Breyer's general concern, but it points to a source of variability very different from that emphasized by Justice Breyer. Contrary to the common view, the problem does not lie in insufficiently clear instructions. The problem instead is the effort to measure attitudes in dollars. Even general and open-ended instructions can produce a high degree of predictability if the response mode is appropriate. Even specific and tailored instructions are likely to produce a high degree of unpredictability if the wrong response mode is used.

To obtain the virtues associated with the rule of law, emphasized by Justice Breyer and many others, the legal system should counteract the arbitrariness that comes from the unbounded scale of dollars. Of course, rule-of-law virtues are not the only virtues of a legal system,<sup>148</sup> and constraints

---

147. See *BMW of N. Am., Inc. v. Gore*, 116 S. Ct. 1589, 1605-09 (1996) (Breyer, J., concurring).

148. See JOSEPH RAZ, *The Rule of Law and Its Virtue*, in *THE AUTHORITY OF LAW* 210, 211 (1979) (stating that the rule of law is a limited ideal and that other values of a legal system include "democracy, justice, equality (before the law or otherwise), human rights of any kind or respect for persons or for the dignity of man").



on erratic awards may not produce all of the necessary reform. We will turn to these issues shortly.<sup>149</sup>

#### 4. *Deterrence and Retribution*

We have noted that some observers think that the purpose of punitive damages is to provide optimal deterrence, and we have referred to a standard economic argument based on this claim. Our findings here strongly support the following conclusion: If optimal deterrence is the purpose of the award of punitive damages, the jury system is an extremely bad institution. This is so for two reasons. The first has to do with the jury's motivations. The second has to do with the jury's capacities.

First, ordinary people do not spontaneously think in terms of optimal deterrence when asked questions about appropriate punishment, and it is very hard to get them to think in these terms. People come to the role of juror with retributive intuitions, and it remains unclear whether and to what extent the courtroom can overcome those intuitions. Perhaps deterrence plays some role in actual awards,<sup>150</sup> and perhaps it would be possible to shift the jury's attention in the direction of deterrence through more insistent and more carefully designed jury instructions.<sup>151</sup> Our study does not rule this possibility out of bounds. But together with other studies that show a jury's reluctance to follow instructions on the purpose of punitive awards,<sup>152</sup> it does give reason to question whether large-scale shifts are likely. Even if focused on deterrence, a jury will be influenced by moral judgments with a retributive dimension, and these judgments will point in the direction of high awards for conduct that is outrageous but likely to be detected (perhaps a murder or an environmental disaster, for example).

---

149. See *infra* Section IV.B.

150. Cass R. Sunstein, *Damages in Sexual Harassment Cases* (1998) (unpublished manuscript, on file with author), finds real world evidence that punitive awards in sexual harassment cases are positively correlated with one and only one factor: evidence of harassment of someone other than the plaintiff.

151. See Polinsky & Shavell, *supra* note 7, at 957-62 (offering model jury instructions). Polinsky and Shavell want the judge to say: "Your principal task is to estimate the likelihood that the defendant might have escaped having to pay for the harm for which he or she should be responsible. . . . You should use the Table below to determine the punitive damages multiplier that corresponds to your estimated probability of escaping liability." *Id.* at 957.

152. See Reid Hastie, David Schkade & John Payne, *A Study of Juror and Jury Judgments in Civil Cases: Deciding Liability for Punitive Damages*, 22 *LAW & HUM. BEHAV.* (forthcoming Aug. 1998) (manuscript at 12-13, on file with the *Yale Law Journal*) (reporting the outcomes of mock jury tests of punitive damages). After deliberations during which the jurors received copies of the judge's instructions, participants were asked specific questions about the important elements on which they were to make judgments about punitive damages. A lenient grading of this written recall comprehension test yielded a median score of only five percent correct. See also Michelle Chernikoff Anderson & Robert MacCoun, *Psychology of Juror Assessments of Compensatory and Punitive Damages* (paper presented at Fifth International Conference on Social Justice Research, Reno, Nevada, June 27, 1995, on file with the *Yale Law Journal*) (showing leakage between compensatory damages and punitive damages).

Second, jurors are not likely to be good at promoting optimal deterrence even if this is what they seek to do. If, for example, punitive damage awards are supposed to be grounded in the probability of escaping detection, it is sufficient to say that ordinary people are very bad at making post hoc probability judgments. In order to assess the probability of detection with any precision, people must master a high degree of technical knowledge about a wide variety of subjects. Hindsight bias will almost inevitably confuse the assessment.<sup>153</sup> A jury is likely to find that a bad outcome was likely to occur if it in fact occurred. Various heuristics and confusion will in all likelihood infect the assessment. If optimal deterrence is the goal, some institution other than a jury—probably an administrative body composed of experts and charged with the specific task—would be much better.

Our findings strongly suggest that the best justification for continued use of the jury involves the desire to elicit, and to make relevant for law, the community's judgment about appropriate retribution. For those who believe that retribution is not a good use of the system of civil and criminal law, this justification will of course be unconvincing. And for those who believe that deterrence is the most important ground for civil law, with retribution playing a modest or supplementary role, a system that elicits and uses community judgments without noise will at best produce a modest improvement.

We therefore arrive at a simple conclusion. To the extent that there is an argument for continued reliance on the jury in awarding punitive damages, it must depend on the possibility of obtaining, in individual cases, an understanding of the public's judgment about the egregiousness of the wrong and the appropriate degree of response.<sup>154</sup> The task is to find a method to obtain that judgment without introducing arbitrariness and noise.

## B. *Punitive Damages Reform*

Punitive damages reform should attempt to ensure that juries are charged with performing tasks that they can perform well. It should also relieve juries from having to perform tasks that they perform poorly, thereby producing excessive unpredictability,<sup>155</sup> confusion, and arbitrariness. It would be

---

153. On the general topic of hindsight bias, see Baruch Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOL. 288 (1975); Rachlinski, *supra* note 2. See also Reid Hastie, David A. Schkade & John W. Payne, *Juror Judgments in Civil Cases: Hindsight Effects on Liability Judgments* 31 tbl.2 (1998) (unpublished manuscript, on file with the *Yale Law Journal*) (showing a hindsight bias directly for punitive damages, with a probability judgment rising from 34% to 59% because of hindsight).

154. See Galanter & Luban, *supra* note 8, at 1433.

155. There are questions as to how detrimental it is for jury judgments to be unpredictable and what kinds of unpredictability might be acceptable. If prospective defendants can assess judgments that are unpredictable in particular cases in order to get a sense of "expected value," unpredictability may not be so bad from the standpoint of optimal deterrence. One reason that unpredictability is undesirable is that it may make planning more difficult if expected value is costly or impossible to calculate, and it may create

reasonable to react to our study by suggesting a simple reform: Juries should decide questions of civil liability, just as they do questions of criminal liability. But judges should decide on the appropriate level of punitive damages, just as they do criminal punishment, subject, in both cases, to guidelines laid down in advance.<sup>156</sup> The result would be a system of *civil sentencing*. Of course, there is a possible problem with judicial determinations of punitive awards, just as in the case of judicial choice of criminal sentences: In both cases, judges are scaling without a modulus, and different judges will reach different conclusions, thus producing arbitrariness. Hence there is good reason for guidelines and constraints on judges. In any case, our study provides strong support for the practice, found in some courts, of reviewing punitive awards to ensure that they are consistent with general outcomes in other cases. Judges need not fear that this practice is anti-populist, for, as we have seen, the award of any particular jury may well fail to reflect the community's sentiment about an appropriate dollar award, even if it accurately reflects the community's sentiment of punitive intent.

To evaluate these and other possible reforms, it is important to distinguish, more carefully than we have thus far, among three possible objections to the idea of using the dollar amounts determined by the juries, as the legal system currently does. The first objection emphasizes sheer unpredictability. The problems here are that potential defendants are not given fair notice, and that similarly situated people are not treated similarly, in large part because any particular jury's judgment about the appropriate dollar award is unlikely to reflect the judgment of any other jury, or of the community as a whole about the appropriate dollar award. The second objection points to defective calibration, that is, to poor translation of punitive intent into dollars. The problem is that juries lack the information that would enable them to undertake good or accurate translations, since ordinary people cannot know the effects of a particular dollar award on a particular class of defendants. The third objection is directed against punitive intent, and points to improper grounds for judgment. Here the complaint is that the jury is either focusing on irrelevant factors, giving undue weight to relevant factors, giving insufficient weight to relevant factors, or ignoring relevant factors. If optimal deterrence is the goal of punitive damages, the outrage model will be quite unappealing. Moreover, public judgments are mediated by social norms, and if those norms are objectionable (as they might be, for example, in the area of sexual harassment),

---

overdeterrence in risk-averse actors; another reason is that unpredictability may be unfair, since people who are similarly situated are not treated similarly. Of course, predictability is not the only value. A modest degree of unpredictability may well be better than a system in which, for example, jury judgments are predictably too low or too high.

156. For various views on the Sentencing Guidelines, see Symposium, *The Federal Sentencing Guidelines: Ten Years Later*, 91 NW. U. L. REV. 1231 (1997).

a noise-free punitive damage judgment based purely on community sentiment is also objectionable.

These three objections lead naturally to three different directions for possible reform. We should stress at the outset that all three reforms allow the jury not to focus on dollars, though for quite different reasons. We offer these reforms partly as thought experiments designed to help specify problems with the current system. To evaluate any of them, and to choose among them and more familiar alternatives, a great deal of additional work would have to be done, much of it involving a comparative analysis of different institutions. What we claim to show is how legal goals mesh, and fail to mesh, with an understanding of the psychological underpinnings of punitive awards.<sup>157</sup>

### 1. *Predictable Populism: Community Judgments About Dollars*

The most modest reform proposal is straightforward. Its goal is to obtain a true understanding of community judgments—true in the sense that it filters out the noise and arbitrariness that come from asking random groups of twelve people to determine the community's judgments about dollar amounts. If this could be done successfully, it would, in one simple stroke, reduce the problem of unpredictability by a large factor (in the illustrative data used here, by a factor of 2.18).<sup>158</sup> The result would be a form of predictable populism.

We have seen that if particular juries are asked to produce dollar awards as indicators of community sentiment, there will be a great deal of variability. There is also a degree of susceptibility to anchors that have little or no normative weight. But if juries are asked to produce not dollar amounts but either punishment ratings or punishment rankings, the number that results can be turned into what we might call "true dollar awards" by the simple step of taking the jury's rating or ranking and using a population-based calibration function to produce a dollar value.

Because this approach does what the current system seeks to do with so much less noise and arbitrariness, it should be counted as a nearly unambiguous improvement. It accepts the sovereignty of community judgments with respect to punitive damages, even dollar awards, and it uses the jury to obtain an estimate of what the population as a whole, if equally informed,

---

157. An important point in this connection is that ordinary people are intuitive retributivists. Moreover, there is a serious question about the role of retribution with respect to corporations and firms. See Polinsky & Shavell, *supra* note 7, at 948-52. The retributive idea is most naturally and simply introduced with respect to people who have imposed harms. Cf. Hampton, *supra* note 8, at 112-22 (discussing whether retribution is a "distinctive and legitimate" form of hatred). The goal is to make wrongdoers suffer. But firms are not persons, and when punitive damages are imposed, the people who are injured, or made to suffer, may not be wrongdoers at all. Thus, a punitive damage award imposed on a firm may well end up injuring not "the firm" so much as consumers, stockholders, employees, and managers who had nothing to do with the underlying wrong. It is far from clear that juries awarding punitive damages are aware of this point, and it is also far from clear that they can be easily convinced that the point is correct.

158. See *supra* Table 2.

would want to have done. The use of a calibration function obtains a more accurate reading of the population's dollar judgment, in a way that eliminates errors introduced by reliance on individual juries for dollar amounts. Through this route it would be possible to produce much more predictability without sacrificing anything else.

a. *Administrative Issues*

There is, to be sure, a serious administrative challenge in generating and using a calibration function. An especially hard question is how to define the category of cases against which any particular case would be assessed.<sup>159</sup> There is also a question whether to alter the calibration function when social norms change. But these difficulties may not be insuperable; at the very least, experimentation along these lines may be worthwhile. A set of common scenarios would be devised in different areas of the law, for which both punishments and awards would be determined.<sup>160</sup> The calibration could be done (say) every five years, to take account of changing social norms.

As Part III above showed, the most important step would not be especially difficult to carry out. Data would also be collected on the effects of firm size and any other pertinent factors. Once this has been done, the procedure in individual cases would be simple. The court would elicit the jury's intention to punish; this, in addition to firm size and other factors, would be used to come up with an estimate of the population's median judgment for that scenario.

b. *Two Phases*

To accomplish this task, the judge would be required to put the jury's focus on intended judgments about punishment rather than dollars; it is the calibration function that would turn those judgments into dollars. Thus a two-phase process would be required. The jury instruction in the first phase might read roughly like this:

Usually, the legal system awards damages in order to compensate the plaintiff for the wrong done by the defendant. These damages are called compensatory damages. But in cases of severe wrongdoing—extreme recklessness or intentional harm—the legal system also allows you to impose a punishment, one that goes beyond the amount

---

159. We have dealt here with personal injury cases, but this may be too narrow or too broad a category. In the context of discrimination, for example, should race and sex discrimination cases be separated? How neatly can contract and tort law be separated, and what subdivisions are appropriate within those categories?

160. Some cautionary notes, in the context of pain and suffering, emerge from Baldus et al., *supra* note 106, at 1125, which discusses possible problems with juries' understanding of cases.

necessary to compensate the plaintiff. The purpose of this punishment is to deter future wrongdoing and to reflect your view about the need not only for compensation but also for punishment, because of the special circumstances of the case. Any punishment will eventually take the form of dollars. Your judgment will be translated into dollars by taking account of the general population's views on how to turn your ranking [or rating] into dollars.

The judge might conclude: "Your choice is very important. It will be the basis for the financial punishment that will ultimately be imposed on the defendant."

For the first phase, it is necessary to make two judgments: about the wording of any questions to be put to the jury, and about the relevant scale on which the jury's normative judgment will be expressed. If a goal is to promote predictability in awards, perhaps the most obvious route would be to require juries to make judgments along a bounded numerical scale. A jury might, for example, be asked to decide where the case falls on a scale of 0 to 6, or 0 to 10.<sup>161</sup> The advantage of this approach is that it would lead to greater predictability with respect to damage awards, and it would thus avoid much of the randomness that characterizes jury selection of dollar amounts. It may therefore be worthwhile for states to experiment with some such approach, at least in some of the settings that call for punitive damages.

But there are problems with a bounded numerical scale. If they are accompanied only by verbal descriptions of the sort we have given here, the relevant numbers are likely to be perceived as highly artificial.<sup>162</sup> In practice, such scales can work quite well; consider, for example, the existence of popular and relatively informative movie ratings that use such scales, or ask whether it might not be possible to rate basketball players, judges, lawyers, or law review articles on such a scale. But juries are likely to be skeptical that the numbers have much meaning. There is a further problem: With a relatively vague scale, different people are free to interpret the labels in different ways. And perhaps most importantly, our data show that there is more (implicit) agreement on rankings than on absolute numbers. Hence the use of numbers will produce less predictability than an attempt to produce rankings.<sup>163</sup>

---

161. Relevant discussion and a related proposal can be found in the Baldus et al. study. *See id.* The authors propose judicial review of awards by reference to comparison cases, building on existing practice. *See id.* at 1141-88. Our proposals represent a more dramatic departure from existing practice.

162. There is also a risk that a small set of numbers will make it difficult for the jury to draw enough distinctions among cases; a large set of numbers will reintroduce some of the problems of an unbounded dollar scale. In some especially heinous cases, the jury may well be drawn to the largest possible number (6 or 10), and it is possible that this will make it harder to draw relevant distinctions.

163. The intuition behind this judgment should be straightforward. If you are asked to assess something—a meal, a conference, a lecture, a concert, a brief, a law review article, a judicial opinion—you may well have a hard time making use of a bounded numerical scale. (On a scale of 0 to 10, how was last night's dinner?) It is much easier to make comparisons—to compare meals, conferences, lectures, briefs, and the like against one another. Comparisons avoid the need to generate uniformity among heterogeneous people with respect to the meanings of the key terms.

It follows that a superior alternative is probably to provide juries not with a bounded numerical scale but with a calibrated set of scenarios of punitive damage cases, set so as to show mild wrongdoing not subject to any award, all the way to extremely egregious wrongdoing requiring a very substantial award.<sup>164</sup> The jury would be asked to assess the case at hand in terms of the sample cases, not with any dollar amounts, nor even with numbers. The sample cases could be ranked in advance, and the jury could be informed of the ranking. Or the jury could be given a set of, say, ten scenarios, provided without any preset ordering (although the scenarios would have been ranked in a large pretest). The jury's task might be to compare the case at hand successively with each of the ten scenarios, to determine whether it is worse or better than those scenarios. The relevant score would be the number of scenarios that are better, plus one. This produces a ranking of the case in the set of eleven, without requiring the jury to rank all eleven. There is no real need for the jury to spend time discussing the details of the orderings of fictitious scenarios. Different scenarios would be provided for different areas of the law: libel, products liability, damages to natural resources, assault, and so forth.<sup>165</sup> The jury would also be allowed to go off the scale, by deciding that the case at hand is less deserving of punishment than the least serious case, or that it deserves more severe punishment than the most serious case.<sup>166</sup>

---

164. See Baldus et al., *supra* note 106, at 1154-60 (providing a detailed and instructive discussion of comparison cases for additur and remittitur judgments); see also Randall R. Bovbjerg et al., *Valuing Life and Limb in Tort: Scheduling "Pain and Suffering,"* 83 NW. U. L. REV. 908 (1989).

Note, however, that a great deal depends on the choice of the relevant comparison set. If the comparison cases are extremely outrageous, it would be possible to lower punitive damages awards in a systematic way. Conversely, if the comparison cases are not so bad, it would be possible to ensure high awards. There is also a possibility of framing effects. The phenomena of "tradeoff contrast" and "extremeness aversion" may well play a large role in jury determination. See Kelman et al., *supra* note 31, at 288. Extremeness aversion refers to the fact that people do not like to take a position that falls on a pole on a continuum; hence people like an option better if it is intermediate. Extremeness aversion can produce "compromise effects," as when people rank an option under consideration in between the poles. *Id.* at 288, 298. Tradeoff contrast arises when the option under consideration is evaluated more favorably in the presence of similar but clearly inferior options than it would otherwise appear in the absence of such options. See *id.* at 288-89. (The option can also be evaluated less favorably in the presence of similar but clearly superior options.) A good way to handle extremeness aversion is with cases that are genuinely extreme—involving, for example, high outrage numbers for genuinely rare and outrageous cases, and low outrage numbers for cases that do not plausibly involve intentional or even reckless wrongdoing.

165. Thus, for example, Baldus et al. discuss certain grounds for creating a typology for purposes of additur or remittitur for medical malpractice, including, in the context of intentional wrongdoing: (1) obstruction of justice, such as destruction of evidence of negligence; (2) other dishonesty, such as failure to disclose to the patient available options; (3) delivery of nonapproved care; and (4) intent to cause harm in the delivery of care. See Baldus et al., *supra* note 106, at 1155 & tbl.10. There is a similar hierarchy for willful disregard. See *id.* at 1155-56 tbl.10. What we are suggesting is that ideas of this kind might be useful sources from which to develop scenarios for jury use.

166. If the no-comparison effect had been stronger, this ranking procedure would have been clearly superior to ratings. The limited effect of the no-comparison condition merely removes one reason for adopting the procedure, but the procedure is still far better than the status quo, and seems better overall than rankings.

A key advantage of this ranking approach, as with ratings, is that the jury would be selecting a dollar amount only implicitly. As compared to a bounded scale, it requires a more concrete and intuitive task and thus may stand as the best way of promoting predictability while maintaining a substantial role for the jury. There are some disadvantages, too. It is possible that the most sensible versions of this approach will offer the legal system's rankings, but the jury may disagree with the rankings, and this may confound the whole enterprise. While the ranking might serve to educate jurors about community values and judgments, some juries might reject what they are told. On this count, ratings have an advantage. Another disadvantage of the ranking approach is that it requires a legislature or commission to generate the scenarios, a more complex and contested task than that involved in using a bounded numerical scale. Our overall judgment is that rankings are likely to work better than ratings; doubtless a fair bit of experimentation would make sense to find what works best. An experimental study, in the context of pain-and-suffering awards, has shown that the practice of offering examples of awards in similar cases can dramatically decrease variability.<sup>167</sup> Our own study of context provides additional support for our proposal to provide a context to cases, and it indicates that the benefits of this practice would extend to both rankings and ratings.

c. *Using Scenarios and Cases: A Note on Existing Practice*

Some aspects of these proposals may seem unusual. Indeed, it has long been impermissible to describe award amounts to juries through actual or hypothetical cases, mostly because of the allegedly prejudicial effect of such references.<sup>168</sup> Moreover, judicial review of punitive damage judgments tends to be highly individualized, and thus it relies on context-independent intuitions (except to the degree that past experience may give judges a sense of context and comparison).<sup>169</sup> Thus, adjustments are often made by reference to the ratio between compensatory and punitive damages (a crude test, because on no theory of punitive awards is that ratio a good way to test the validity of the punitive award) and ad hoc judgments about what seems "shocking."<sup>170</sup>

The idea of examining comparison cases is not, however, entirely foreign to the legal culture. For instance, damage schedules and scaling through

---

167. See Saks et al., *supra* note 140, at 250-52 & tbl.III. Of course, there is a risk of strategic behavior on the jury's part if the jury becomes aware of the relevant conversion function, and steps must be taken, perhaps through jury instructions, to counter this risk if it materializes.

168. See D.C. Barrett, Annotation, *Propriety and Prejudicial Effect of Reference by Counsel in Civil Case to Amount of Verdict in Similar Cases*, 15 A.L.R.3d 1144 (1967).

169. See Baldus et al., *supra* note 106, at 1132-33.

170. See, e.g., *Malandris v. Merrill Lynch, Pierce, Fenner & Smith, Inc.*, 703 F.2d 1152, 1177 (10th Cir. 1981) (reviewing an award without engaging in comparisons); Baldus et al., *supra* note 106, at 1133.



examples have been used successfully in the settlement of mass tort cases.<sup>171</sup> Such ideas have often been discussed in the context of pain-and-suffering awards.<sup>172</sup> In a prominent case involving such awards, Judge Kearse, writing for the Second Circuit, carefully compared the case at hand with fourteen other cases.<sup>173</sup> Such comparisons have also been made in the context of punitive damages, at least in the process of judicial review.<sup>174</sup> In an influential case, the U.S. Court of Appeals for the Ninth Circuit said that a district court should "look to awards in similar cases and to its own experiences" in order to arrive at "a figure derived from the facts of the case at hand."<sup>175</sup> Of course, the Supreme Court has not insisted on this requirement as a matter of constitutional law. In an echo of some early death penalty cases,<sup>176</sup> the Court said that each case might be taken as *sui generis*: It emphasized that punitive damage judgments require juries to "make a qualitative assessment based on a host of facts and circumstances unique to the particular case before it. Because no two cases are truly identical, meaningful comparisons of such awards are difficult to make."<sup>177</sup> The Court said that a comparative approach cannot be a "'test' for assessing the constitutionality of punitive damage awards," even though it would not "rule out the possibility that the fact that an award is significantly larger than those in apparently similar circumstances" might be relevant to the constitutional issue.<sup>178</sup>

The Court's refusal to impose a constitutional requirement of comparing cases is understandable in light of principles of federalism and gaps in the Court's knowledge of the actual world of punitive damage awards. Nothing said here demonstrates that such a requirement must be mandated as a matter of constitutional law. But it is clear that a principal concern with the existing system stems from inadequate constraints on jury discretion, and it is now

171. See Francis E. McGovern, *The Alabama DDT Settlement Fund*, LAW & CONTEMP. PROBS., Autumn 1990, at 61, 67, 73-74; Francis E. McGovern, *Resolving Mature Mass Tort Litigation*, 69 B.U. L. REV. 659, 682 (1989).

172. See *Geressy v. Digital Equip. Corp.*, 980 F. Supp. 640, 656-60 (E.D.N.Y. 1997); *Baldus et al.*, *supra* note 106, at 1123-25, 1134-40; *Bovbjerg et al.*, *supra* note 164.

173. See *Martell v. Boardwalk Enters., Inc.*, 748 F.2d 740, 752-53, 755 (2d Cir. 1984).

174. See, e.g., *Bogan v. Stroud*, 958 F.2d 180, 186 (7th Cir. 1992); *Estate of Korf v. A.O. Smith Harvestore Prods., Inc.*, 917 F.2d 480, 485 (10th Cir. 1990); *Ismail v. Cohen*, 899 F.2d 183, 186-87 (2d Cir. 1990); *Schultz v. Thomas*, 649 F. Supp. 620, 624-25 (E.D. Wis. 1986); *Sherrod v. Piedmont Aviation, Inc.*, 516 F. Supp. 46, 56 (E.D. Tenn. 1978); cases cited *supra* note 105.

175. *Morgan v. Woessner*, 997 F.2d 1244, 1257 (9th Cir. 1993).

176. See, e.g., *Woodson v. North Carolina*, 428 U.S. 280, 303-04 (1976) (plurality opinion) (invalidating a mandatory death penalty statute on the ground that it did not allow individualized consideration).

177. *TXO Prod. Corp. v. Alliance Resources Corp.*, 509 U.S. 443, 457 (1993) (plurality opinion).

178. *Id.* at 458. Concurring, Justice Kennedy pointed to the high likelihood of legitimate inconsistency in jury results. In his view, this is partly a function of the fact that a jury is empaneled in a single case, not as a permanent body, and partly a function of the generality of jury instructions. See *id.* at 468 (Kennedy, J., concurring). Thus a lower court suggested that review of other juries' determinations "would undercut the jury system" and that apparently inconsistent awards are products of the fact that "juries hear unique facts and are given dissimilar instructions." *In re the Exxon Valdez*, No. A89-0095-CV (HRH), 1995 U.S. Dist. LEXIS 12952, at \*7 n.7 (D. Alaska Jan. 27, 1995).

clear that a serious problem arises from the response mode of dollars and the difficulty of generating predictable dollar amounts. Our findings support the wisdom of the appellate practice of comparing punitive damage awards, and they suggest the possibility that dollar awards not based on comparisons amount to stabs in the dark and may be constitutionally infirm.

## 2. *Technocratic Populism: Using Punitive Intent but Not Dollars*

The second kind of reform we have suggested would attempt to elicit the jury's punitive intent, or its judgment about appropriate punishment, but would not ask the community to make decisions about dollar amounts. On this view, it is agreed that the jury's intention to punish is what should govern punitive damage awards. In this way, the outrage model is accepted on normative grounds. The problem is that the most modest reform proposal, just described, perpetuates the crucial defect of the current system; that is, it relies on the ability of ordinary citizens, and hence the community, to translate punitive intention into dollars (which, we have argued, results largely in stabs in the dark). There is an analogy here with the criminal justice system as it now stands: Juries make decisions about criminal liability (decisions that undoubtedly have a dimension of "punitive intent"), but judges, within the constraints of applicable guidelines, make decisions about sentencing, presumably because of their greater expertise and insulation from irrelevant or illegitimate factors.

Here is a simple argument for the second kind of reform. People are unlikely to know what it takes to hurt different people of different means through financial punishment. They certainly do not know what it takes to hurt an organization. Whether or not an organization is the defendant, juries using unbounded dollar scales lack a modulus that would give meaning to their estimates of different magnitudes. Thus, a conversion formula might depend on the community's normative judgments about intended punishment, which are not only predictable but also worth using, rather than on community judgments about dollar amounts, which can be made more predictable but which may not be worth using. To make this work well, experts would help develop the conversion formula. Hence, the legal system might endorse a form of *technocratic populism*, in which popular judgments are the foundation for expert determinations. Here are the two phases of a possible reform.

### a. *Phase One: The Jury's Role*

If the legal system is interested in the jury's punitive intent, but not in the community's judgments about dollar awards, phase one would look very much

like phase one in the modest recommendation described above.<sup>179</sup> The goal here too would be to obtain the jury's judgment about ratings or rankings. The judge's instruction might be similar to that described above, but with a different ending, such as, "It is the job of the judge and the legislature, and not the jury, to decide on specific dollar amounts. Your job is to help in that task by informing the court of how severely, in your view, the defendant deserves to be punished." Here too there is a good argument that rankings are preferable to ratings, because they are less artificial and do not run up against a judgment that the numbers on a bounded scale are arbitrary and meaningless.

b. *Calibration and Phase Two: Translating Jury Judgments into Dollars*

In designing phase two, we assume that the jury's judgment about appropriate punishment is the appropriate foundation for the award, and that the problem for correction is that a jury is in a poor position to know what the effects of dollar awards will be. A good conversion formula must ensure that the expressed punitive intention is turned into awards that are both predictable and faithful to what the jury truly intends to accomplish. The translating institution would have to know a great deal about the effects of dollar awards on both individuals and firms. It would also have to make judgments about the effects of the defendant's income and wealth. Of course, the judgment about how to translate would involve many evaluative judgments. A legislature or commission charged with making those judgments might ask, for example, about the effects of various kinds of awards on both individuals and firms: When high punitive awards are given, how exactly do firms suffer? Are those who suffer high-level officials, clerical workers, or consumers? Judgments about dollar awards would emerge through engagement with such issues.

3. *Civil Sentencing and Bureaucratic Rationality:  
The Partial or Complete Elimination of Juries*

Our findings put in sharp relief a large and almost entirely unexplored problem: whether, in light of what is or might be known about human psychology and cognition, lay people are willing and able to make judgments in the way that the legal system deems desirable. There is good reason to believe, for example, that if punitive damages are designed to produce optimal deterrence, juries should be eliminated, for it is doubtful that they can be made to carry out that task. If optimal deterrence is the goal of punitive damages, it would probably be better for the judgment to be made by a judge or (better still) by a specialized regulatory agency.

---

179. See *supra* Subsection IV.B.1.b.

Under the most natural justification for allowing punitive damages to be awarded by juries, punitive awards should be seen as serving a retributive or expressive function, in which social norms are brought to bear on certain behavior, thereby offering a public “statement” about appropriate outrage and punishment. There can be little doubt that judgments about outrage and punishment may diverge from judgments about optimal deterrence; people would surely want to punish someone who engaged in extremely outrageous conduct via punitive damages even if the probability of detection were 100%. Our discussion thus far suggests a simple idea: If juries are appropriately used, it is because their intuitions are more or less acceptable, assuming that they can be adequately disciplined. We know a good deal about these intuitions. They exhibit high sensitivity to outrage (which probably implies low sensitivity to detection probability<sup>180</sup>), substantial sensitivity to harm (which may not be fully expressed in the amount of compensatory damages), great sensitivity to firm size, and a backward-looking, retributive focus.

A decision to retain the jury system implies that these intuitions are appropriately used for purposes of civil punishment, at least as a good first approximation. But if our diagnosis of the intuitions is correct, there is a risk in a translation formula: Juries may feel exposed to the possibility of being ignored and may respond by behaving strategically. There is a limit to the extent to which the translation formula can stray from common intuitions.

The most serious objection to our second proposal comes from the view that juries do not and cannot easily be made to base their decisions on the proper grounds.<sup>181</sup> On this view, the problem is that a jury’s intention to punish, even if well translated, is not an adequate way to assess punitive damage awards, because it is, from the normative point of view, too sensitive to irrelevant factors and too insensitive to relevant factors. In the criminal law, juries are not asked to identify their punitive intent, with a separate calibration by the judge. Instead the determination is made by the judge subject to the Sentencing Guidelines. Why would not the same approach make sense for punitive awards?

The claim that juries cannot be used to promote the aims, properly understood, of the system of punitive damages might be based on at least three possible objections. The first is that retributive or expressive goals would be better carried out via the criminal law, or even by regulatory law at the local, state, and national levels. Certainly in the environmental area, expressive and

---

180. *But see* Sunstein, *supra* note 150, at 15 (finding that the existence of other harassment incidents is correlated with higher punitive awards).

181. *See* Hastie et al., *supra* note 152 (manuscript at 34). In that study, two-thirds of mock juries that deliberated to a verdict decided that punitive damages were warranted in actual cases in which appellate and trial judges had concluded otherwise. *See id.* (manuscript at 44 tbl.4). This strong tendency to find liability for punitive damages was partly a product of the jurors’ failure to consider the judge’s instructions; juries that discussed those instructions more fully were less likely to impose punitive damage liability. *See id.* (manuscript at 50-51 figs.2-3).

retributive motivations have played a large role in areas not involving juries. Juries awarding punitive damages for purposes of retribution may not be sufficiently informed by relevant facts or capable of reliably expressing relevant values. Thus juries may not serve as a sensible complement to the remaining fabric of the law. Even if we focus on retributive goals, punitive damages may not fit with the rest of the law in a way that creates a sensible regime of retribution.

The second objection is that deterrence is the appropriate goal of punitive damages and that juries cannot be made willing and able to inquire reliably into how to achieve optimal deterrence. It appears that outrage is inattentive, or insufficiently attentive, to factors that are central to the goal of optimal deterrence, most prominently the probability of detection. If this is so, then an administrative or regulatory body would better achieve that goal. Certainly an inquiry into the probability of detection seems like a factual one that juries lack the information to undertake properly. And if outrage is attentive to irrelevant factors, the case for abandoning the jury is strong, even or, perhaps, especially when some form of punishment is desirable. Predictability is obtainable, as suggested in our earlier reform proposals, but perhaps without satisfying the substantive goals of a well-functioning system of punishment.

The third objection is that existing community judgments about intended punishment should, in some contexts, be "impeached" by legislative deliberation. The community's sentiment might depend, for example, on social norms that the legal system should not recognize. It might be thought that in some cases, the community's judgments are too lenient or too severe. In the area of racial discrimination or sexual harassment, for instance, a legislative body might question what juries are likely to do. Perhaps juries would be insufficiently sympathetic to plaintiffs who suffer from injuries not fully recognized as such by existing social norms. Juries are emphatically populist institutions; in a sense, our study's basic point has been to give more precise content to this commonplace belief. In a nonpopulist republic that sometimes distrusts community norms, displacement of general sentiment is hardly rare.

If we agree that the intention to punish is relevant to the punitive award, but not decisive, there is a simple response: Use the jury's intention to punish as one among a set of factors for judicial consideration in imposing punitive damage awards. The judge might be required to consider as well the size of the defendant, the probability of detection, the illicit character of the defendant's gains, and other factors. A general requirement of this kind would not impose enormous demands on the legislature involved in punitive damages reform. As just stated, however, a risk with such an approach arises from the fact that weights have not been given to the various factors; the absence of weights raises the danger that judicial determinations will also suffer from unpredictability. Hence, the legislature or commission might attempt to give

greater guidance by, for example, offering scenarios accompanied by dollar awards, creating ranges, or providing floors and ceilings.

Certainly more dramatic alternatives can be imagined, including those that dispense with a jury entirely. Here the goal might be a form of *bureaucratic rationality*. An administratively operated schedule of fines and penalties seems better than juries at producing deterrence, for such a system would reduce the costs of decision and probably reduce the costs of error as well. There are many analogies. Discussion of pain-and-suffering awards has included “technocratic” suggestions substantially designed to reduce the jury’s role in the interest of more consistent and more expert judgments.<sup>182</sup> In the context of damages to natural resources, Richard Stewart has suggested that contingent valuation should be replaced by a schedule of damages based on categories of harm;<sup>183</sup> in this way, an antecedent set of administrative or legislative judgments would form the backdrop for judgments by a trustee, thus making it unnecessary to ask what may be hopelessly uninformative questions of individuals about their willingness to pay.

There are real world precedents for this kind of approach in many domains of law, in which ad hoc determinations have been replaced with a system designed to produce more in the way of coherence and rationality. The system of workers’ compensation was created partly because of the high decision costs and randomness produced by case-by-case jury judgments about, for example, the value of a lost limb.<sup>184</sup> In its current form, workers’ compensation attempts to deal with problems of valuation by placing a fixed dollar value on various injuries through a predetermined schedule produced by a legislature or administrative agency.<sup>185</sup> The problem with the workers’ compensation system is its crudeness; it ignores possibly relevant individual variations. Its virtue lies in its speed, inexpensiveness, predictability, and consistency.<sup>186</sup>

In a related but more recent shift, the process of case-by-case judgment with respect to sentencing has been replaced by the more standardized Sentencing Guidelines. A central goal of the guidelines is to discipline the process of “mapping” complex normative judgments onto a relatively less bounded scale of criminal punishments. The fact that judges, rather than juries, have traditionally made decisions about appropriate sentences raises a question

---

182. See Baldus et al., *supra* note 106, at 1125-31; Bovbjerg et al., *supra* note 164, at 939-49.

183. See Richard B. Stewart, *Liability for Natural Resource Injury: Beyond Tort*, in *ANALYZING SUPERFUND: ECONOMICS, SCIENCE, AND LAW* 219, 241-44 (Richard L. Revesz & Richard B. Stewart eds., 1995); see also Murray B. Rutherford et al., *Assessing Environmental Losses: Judgments of Importance and Damage Schedules*, 22 *HARV. ENVTL. L. REV.* 51 (1998) (advocating a damage schedule for environmental losses that is derived from pairwise comparisons of the relative importance of different adverse environmental outcomes or events).

184. See, e.g., RICHARD A. EPSTEIN, *CASES AND MATERIALS ON TORTS* 913-38 (5th ed. 1990).

185. See *id.* at 935-38.

186. See *id.*

about why a similar course is not followed for punitive damages.<sup>187</sup> The basic point is that with both workers' compensation and the Sentencing Guidelines, a process of bureaucratic rationalization has replaced one of relatively ad hoc judgments. It is easy to imagine a similar development with punitive damages.

An even more relevant model can be found in the "grid" used for social security disability determinations, which uses age, educational attainment, and residual functional capacity to produce standardized judgments about disability.<sup>188</sup> Administrative law judges are asked to make case-specific judgments, which become part of an assessment governed by the rule-like grid. It is useful to ask, as a thought experiment, why juries, rather than grid-governed administrative law judges, are not asked to make disability determinations. Surely the answer is that jury determinations would suffer from a range of problems, including insufficient specialization and expertise, inevitable arbitrariness and unpredictability, and confusion stemming from the use of an unfamiliar scale. But if jury judgments would be inappropriate for disability determinations, why do they make sense for punitive damages, or for that matter for judgments about compensation in cases involving pain and suffering or libel? Any answer would have to refer to the legitimate domain of populism in law.

Radical changes of the kind just discussed call for a comparison of the likely performance of different governmental institutions. Dramatic changes might be criticized on the ground that the populist elements of jury assessment should be retained to ensure that public outrage plays a significant role in the legal system. In addition, technocratic substitutes for the jury might be subject to pathologies of their own, perhaps because of their own biases, and because of the pressures likely to be imposed by well-organized private groups. If so, the more modest reforms are better. We have sought to undertake the first step toward that evaluation—specifying the underlying considerations with a better understanding of what produces punitive awards. The disadvantage of the most extreme departures from the current system is also their advantage. They would not rely on the jury's normative judgment about outrageousness and intended punishment. What we have proposed is that there are ways to retain this goal while also diminishing the unpredictability of punitive damage awards.

---

187. There are, of course, differences between the two settings. The distinctive stigma associated with criminal punishment may make it seem especially important to insulate judgments from the passion and zeal that might affect a jury. That stigma may also make specialized experience, and even guidelines, especially important. And juries do, of course, have some control over sentencing through their choices among different theories of criminal liability.

188. See *Heckler v. Campbell*, 461 U.S. 458 (1983); JERRY L. MASHAW, *BUREAUCRATIC JUSTICE* 114-23 (1983).

#### 4. *Evaluating the Alternatives: The Proposed Reforms, Caps, and Multipliers*

It is possible to imagine mixed approaches, drawing on different aspects of our proposals. For example, a jury might be provided with a preselected set of exemplar cases, accompanied by the damages actually or reasonably awarded in these cases. The jury's task might be to assess damages by comparing the case at hand to the preselected cases. This approach would not remove the whole subject of dollar awards from jurors; instead, it would attempt to root punitive awards in a set of comparable antecedent judgments. The damages in the exemplar cases might be based on actual past judgments, on judgments of mock juries, or on judgments of experts in the particular area.

Other mixed approaches might attempt to supply a kind of modulus. Juries might, for example, be given average dollar awards for the type of injury at issue, or intervals (showing where a certain percentage of awards for similar injuries fell), or both average dollar awards and intervals.<sup>189</sup> Doubtless some experimentation would help show which approach works best.

Currently, both the federal government and the states are discussing more conventional reforms, which would impose caps, require punitive damages to be within some multiple of compensatory damages, or allow judges to have a larger role in disciplining jury awards.<sup>190</sup> We might compare various reform proposals by asking which would impose the lowest sum of decision costs and error costs, recognizing that there is no simple metric for assessing these kinds of "costs." It is clear that the chief advantage of caps and multipliers is their simplicity and low administrative cost; their chief disadvantage is that they are unlikely to do much to decrease error costs, and they may even increase them.

If the problem is that juries are not now required to think in terms of optimal deterrence, both caps and multipliers are extremely crude. There is no reason to believe that either of these reforms would ensure that punitive awards are tailored to compensate for the likelihood that injured parties will not bring suit. If optimal deterrence is the goal, the best solution would be to abandon the jury and to delegate power to an institution willing and able to think well about optimal deterrence. If, on the other hand, the problem is that isolated juries generate arbitrary or unpredictable outcomes, and the purpose of reform is to ensure that punitive damage awards capture either the community's intentions with respect to dollar awards or the community's judgments with

---

189. See Saks et al., *supra* note 140, at 243 (presenting a study following these alternative approaches).

190. See *supra* notes 109-110 and accompanying text.



respect to punitive intent, the conventional proposals are also likely inferior to those we have discussed here.

A cap, a multiplier, or judicial oversight will obviously do little to ensure expression of community will with respect to either dollars or punitive intent. A cap has one important advantage: It is easily administered, and it may prevent awards that are plainly excessive. But a cap is, of course, crudely tailored, and it may even increase variability, partly because the cap may serve to anchor jury judgments and thus draw jurors to the upper bound.<sup>191</sup> The best that can be said about a cap is that it will eliminate the most egregiously large judgments at low administrative cost, but this virtue comes with many vices.

A damage multiplier might work somewhat better, in the sense that it would also have low decision costs while allowing more flexibility than a cap. A damage multiplier would permit very high awards when the compensatory damages are especially serious, while preventing a jury from imposing unreasonable punishments on an individual or a corporation. But multipliers are still quite crude, and no theory of punitive damages justifies a multiplier approach. The deterrence theory is not of course intended to secure parity with the compensatory award; when the probability of detection is very low, a high ratio makes sense. From the standpoint of retribution, the compensatory award may not be especially important in cases in which outrageous conduct (firing a gun into a crowd) happened to produce little harm.

Of the proposed reforms, an incremental shift from jury to judicial determinations of punitive damages appears to be the most promising. Such a shift may well produce improvements over the current system, at least if illegitimate or irrelevant factors play too great a role in jury determinations, or if judges, because of their experience with multiple cases, can make more informed judgments. (Again, compare criminal sentencing, undertaken by judges rather than juries.) Judges do reduce punitive awards that appear excessive, and because of their experience judges have many comparative advantages;<sup>192</sup> it is likely that a more general shift toward judicial control with reference to comparison cases would produce improvements. The strongest argument against this shift would be either populist or technocratic: Judges are not likely to be able to capture the community's sentiments with respect to either dollar awards or punitive intent, and judges are not experts on the topic of optimal deterrence. Hence expert guidelines

---

191. See Saks et al., *supra* note 140, at 245-46 (finding that caps favor "those with lesser injuries while further depriving the most seriously injured"); see also Verlin B. Hinsz & Knstin E. Indahl, *Assimilation to Anchors for Damage Awards in a Mock Civil Trial*, 25 J. APPLIED SOC. PSYCHOL. 991, 1016 (1995) (finding an anchoring effect from caps). A cap that the jury is not informed about might produce less variability, but it is not clear that it should be kept secret.

192. See *supra* text accompanying note 169.

subject to democratic checks, and with a more detailed understanding of regulatory goals and instruments, should probably discipline the judge's inquiry.

The most important point is that judges too are likely to have difficulty in mapping normative judgments onto dollar amounts, and while judicial judgments may reduce variance, there is likely to be a continuing problem of erratic judgments or the use of anchors that introduce arbitrariness of their own. Judicially assessed punitive awards might well replicate some of the problems with judicially determined sentences. Any movement from jury to judicial control of punitive damage awards might well be accompanied by some form of scaling or scheduling, perhaps building on the old practices of *additur* and *remittitur* in a way that is psychologically well informed. As we have suggested, various reform combinations and alternatives might be imagined, including dollar awards that are chosen after exposure to comparison cases.<sup>193</sup>

What about the problem of interest-group dominance of any legislative reforms? There can be no doubt that interest-group pressures will play a role in any effort at punitive damage reforms, and if some such reforms are especially likely to be immune from such pressures, there is more reason to consider those particular reforms seriously. Of course damage caps may be a product of pressures from well-organized groups seeking a simple "fix" for the existing problem. Supposedly expert determinations may also be much affected by pressures of this kind, a fact that raises questions about both technocratic populism and bureaucratic rationality. Of the proposals discussed here, predictable populism and civil sentencing may be the least subject to interest-group maneuvering—the first because the goal of this kind of reform is so straightforward, the second because judges are insulated from this kind of political pressure. In the abstract, however, judgments about the role of interest groups in future reforms seems highly speculative.

We do not have sufficient information to evaluate all the possible alternatives here. But we can offer two general conclusions. First, there is a great deal to be said for the incremental step of civil sentencing, building on current practice and ensuring, in every jurisdiction, a serious oversight role for judges, calling not for individual judicial judgments about individual cases, but for judicial comparisons among various similar cases, so as to ensure against dramatic outliers.<sup>194</sup> This incremental step would produce some of the gains sought by the first reform proposal discussed above; our findings show why it is less objectionable on populist grounds than it might

---

193. See generally Baldus et al., *supra* note 106 (presenting a study assessing methods for implementing *additur-remittitur* review).

194. Compare Judge Weinstein's fascinating discussion of the use of comparison cases in the context of pain and suffering in *Geressy v. Digital Equipment Corp.*, 980 F. Supp. 640, 657-60 (E.D.N.Y. 1997).

seem. Second, the ideal system of punitive damage awards would not involve juries or even judges, but specialists in the subject matter at hand, who are able to create clear guidelines for punitive awards. These guidelines would be laid down in advance and be based on a firm understanding of different forms of wrongdoing and of the consequences for defendants of different awards. Of course, specialists would need to make several judgments of value, and those judgments should be subject to democratic control. The practical question is whether it is possible to design that ideal system. Experiments in this direction can be found in the workers' compensation system and in the system of administrative penalties and fines.<sup>195</sup>

Consider, by way of summary, Table 4. Our overall evaluations must be tentative, because much depends on some unanswered empirical questions about the likely operation of the different systems.

How, if at all, does what we have said bear on the constitutional analysis of punitive damage awards, an issue of fresh importance in the aftermath of *BMW*?<sup>196</sup> We have identified a source of those jury judgments that may be both unconstitutionally excessive and unconstitutionally arbitrary. When a particular jury's judgment is extreme, an underlying reason may well be the difficulty faced by the jury in using an unbounded scale. Our findings certainly do not resolve the constitutional issue, which depends on the appropriate approach to the Due Process Clause. They do help, however, to fortify the view that some awards should be considered unacceptably arbitrary. High jury awards, requiring payment of some amount \$X, do not necessarily reflect a well-considered community judgment in favor of \$X, nor do jurors match the particulars of cases to dollar awards. Indeed, our findings demonstrate that the populist credentials of any particular jury award may be overstated. Any particular jury's award may poorly measure the judgment of the community as a whole. Moreover, this problem cannot be cured by more detailed instructions from the court, at least if those instructions do not solve the problem of scaling without a modulus.

Of course, the Supreme Court cannot by itself require one of our three proposals or some variant. The selection of reform methods is for legislatures rather than courts. The judicial role should come at the subconstitutional level, through review of punitive damage awards corresponding to the standards outlined above. But the availability of these routes suggests the possibility of retaining a significant role for the jury without providing so large a risk of arbitrariness.

---

195. See *supra* text accompanying notes 184-185.

196. *BMW of N. Am., Inc. v. Gore*, 116 S. Ct. 1589 (1996).

TABLE 4. PUNITIVE DAMAGE REFORM POSSIBILITIES

	DESCRIPTION	ANALOGIES	VIRTUES	VICES	OVERALL EVALUATION
1. CURRENT SYSTEM	ad hoc judgments about particular events, with a populist aspiration	"khadi justice" (as discussed by Max Weber)	allows a role for popular convictions	unpredictability; susceptibility to arbitrary anchors; jury ignorance about effects of dollar awards	hard to defend from any standpoint; the best that can be said is that the current problems are not so serious
2. CAPS	prevents the most excessive awards	current civil rights statutes, which also impose a cap	easy to administer; would prevent egregiously large awards	crudely tailored to any view of the problem or the purpose of punitive awards; may increase variability	unlikely to make things much better and may in some ways make things worse
3. DAMAGE MULTIPLIERS	ties punitive awards to compensatory awards	Sherman Antitrust Act	more flexible than caps; also easy to administer	crude, since compensatory award is a rough guide to appropriate punitive award	a little better than caps, but not much better
4. PROVIDE JURIES WITH OTHER CASES AND THEIR ACCOMPANYING PUNITIVE AWARDS, IN DOLLARS	retains jury authority over dollars	proposals in the area of pain-and-suffering awards	should improve predictability and increase rationality	unclear how jury will respond to prior awards if it disagrees with them; unclear how to make sure the prior awards contain the right amounts	promising way of providing context and cabining judgment, but may be too complex
5. STRENGTHENED OR EXCLUSIVE JUDICIAL CONTROL	"civil sentencing" model	current system of criminal justice	may reduce unpredictability; may produce more overall rationality	may also involve scaling without a modulus; eliminates or reduces populist elements; need to ensure judicial comparison of cases	modest but quite promising response to problems of unpredictability and irrationality
6. POPULATION-WIDE CALIBRATION FUNCTION	a form of "predictable populism"	none	reduces variability; increases predictability	complicated administrative task	intriguing, but possibly too complex and novel
7. EXPERT CALIBRATION FUNCTION	a form of "technocratic populism"	none	solves problem of jury ignorance about effects of punitive awards while preserving centrality of punitive intent	might not trust punitive intent either of juries or the experts	intriguing and promising, but less feasible than reform (5) above and probably less promising than reform (8) below
8. ADMINISTRATIVE PENALTIES	bureaucratic rationality	workers' compensation system; social security "grid"; administrative penalties and fines	could produce both predictability and rationality	experts may not be trustworthy	in principle, the most promising of all, as a way of reducing decision costs and error costs

## V. IMPLICATIONS, EXTENSIONS, AND SPECULATIONS

Our central finding—about the difficulty of mapping relatively shared and stable normative judgments onto an unbounded dollar scale—is relevant to a number of issues of law and policy. If psychology shows that in certain settings people cannot do, or will refuse to do, some or all of what the legal system wants them to do, the normative goals of law will be systematically frustrated. It then becomes necessary to disentangle—as we have attempted to do here—the possible problems with the existing system of jury determinations. It may emerge, for example, that a single jury deviates erratically from population-wide judgments, or that those judgments are systemically flawed because they require judgments on intent to be mapped onto dollars or because they run counter to the basic goals of the legal system.

In many areas, an analysis of the basic sort presented here may well apply. The psychology of various types of damage awards and the goals of the legal system are often in tension. We outline a few possible areas for future investigation and offer a brief note on contingent valuation, which raises overlapping but somewhat different issues.

### A. *Difficult Damage Determinations*

Many damage determinations require juries to undertake magnitude scaling without a modulus, and to do so in settings that lack clear market measures. We offer several examples from the law of compensatory damages, concluding with some general remarks about the relationship between punitive and compensatory damages in terms of our discussion here. An underlying question concerns the appropriate role of normative judgments in settling on the apparently but (as we shall see) controversially “factual” question of what amount would provide “compensation.” This question raises serious issues about the populist and technocratic dimensions of compensatory awards in these domains of the law.

#### 1. *Pain and Suffering*

Awards for pain and suffering raise many of the same questions as punitive damages. To be sure, such awards are nominally compensatory rather than punitive; they ask the jury to uncover a “fact.” But they also involve goods that are not directly traded on markets and require a jury to turn into dollars a set of judgments that are hard to monetize.<sup>197</sup> Judgments about pain

---

197. Thus a standard jury instruction says:

[T]he law allows you to award to plaintiff a sum that will reasonably compensate him for any past physical pain . . . .

and suffering require juries to make a decision about harm (with a likely ingredient, in practice, of intended punishment) and to map that judgment onto a dollar scale. In the absence of uncontroversial market measures to make the mapping reliable,<sup>198</sup> the resulting verdicts are notoriously variable, in a way that raises questions very much like those in the punitive damage setting.<sup>199</sup> In particular, people with similar injuries are often awarded very different amounts of damages.<sup>200</sup> Studies have found that plaintiffs with relatively small losses tend to be overcompensated and that those with large losses tend to be undercompensated and that there is a significant degree of randomness as well.<sup>201</sup> Our study suggests one of the sources of the variability. A judgment about harm, perhaps made in a predictable way on a bounded numerical scale, becomes unpredictable and arbitrary when translated into an unbounded dollar scale lacking a modulus.

In the context of pain-and-suffering awards, anchors appear to be especially important, even if they carry an arbitrariness of their own. Some jurors appear to split the difference between the figures suggested by the plaintiff and the defendant, whereas others use some (fairly random) multiple of medical expenses, and still others fasten on to other aspects of the case as anchors.<sup>202</sup> One study suggests that severity of injury explains only forty percent of the variation in awards.<sup>203</sup>

---

There are no objective guidelines by which you can measure the money equivalent of this element of injury; the only real measuring stick, if it can be so described, is your collective enlightened conscience. You should consider all the evidence bearing on the nature of the injuries, the certainty of future pain, the severity and the likely duration thereof.

GRAHAM DOUTHWAITE, *JURY INSTRUCTIONS ON DAMAGES IN TORT ACTIONS* § 6-17, at 274 (2d ed. 1988). An instruction of this kind offers little more guidance than a typical punitive damage instruction. What is the psychological process by which juries construct such awards? Can juries or judges make predictable or otherwise sensible judgments about dollar amounts? What are the ingredients of those judgments, whatever the instructions say? Most importantly, what exactly does "compensate" mean?

198. An effort at disciplining decisionmaking is made in VISCUSI, *supra* note 27, at 113-16.

199. See U.S. GEN. ACCOUNTING OFFICE, *MEDICAL MALPRACTICE: CHARACTERISTICS OF CLAIMS CLOSED IN 1984*, at 2-3, 18-19, 23-24, 40-42 (1987); Bovbjerg et al., *supra* note 164, at 919-28; Mark Geistfeld, *Placing a Price on Pain and Suffering: A Method for Helping Juries Determine Tort Damages for Nonmonetary Injuries*, 83 CAL. L. REV. 773 (1995); W. Kip Viscusi, *Pain and Suffering in Product Liability Cases: Systematic Compensation or Capricious Awards?*, 8 INT'L REV. L. & ECON. 203, 204-08, 214-19 (1988).

200. See Geistfeld, *supra* note 199, at 784; Saks et al., *supra* note 140, at 243-44.

201. See, e.g., Saks et al., *supra* note 140, at 245.

202. See *id.*

203. See Bovbjerg et al., *supra* note 164, at 923. A recent study shows an additional point: Judgments about pain and suffering are highly sensitive to framing effects. See McCaffery et al., *supra* note 31. In particular, they are sensitive to the endowment effect—the fact that people are willing to pay less to purchase a good than they must pay in order to keep the very same good if it has been initially allocated to them. Thus losses are disvalued more than gains are valued. In the context of pain-and-suffering awards, the question is whether the plaintiff should be entitled to (1) the amount that he would have to be paid, before the fact, to allow the relevant pain and suffering to occur (his selling price); or instead (2) the amount that he would be willing to pay, after the fact, to restore his health to its previous place (his "make whole" price). The study shows substantial differences between (1) and (2), and thus suggests that people are highly subject to framing effects in assessing appropriate awards for damages. Among one group, the "selling price" award was about double the "make whole" award. See *id.* at 1372-73.

Like punitive damages, pain-and-suffering awards are made in a no-comparison condition, and hence juries may fail to provide the kinds of distinctions that would emerge if a set of cases were offered at the time of decisions. And although pain-and-suffering awards are essentially compensatory, there can be little doubt that such awards sometimes reflect jury judgments about the egregiousness of the defendant's behavior. Hence such judgments are likely to have a punitive component.

Much work remains to be done in disaggregating the factors that produce large or small awards for pain and suffering. If the psychology of such awards is similar to that of punitive damage awards, however, it will make sense to consider reforms such as implementing a damage schedule to cabin the jury's judgment<sup>204</sup> or using a set of comparison cases for jury or judicial guidance. The choice among the relevant possibilities depends on a judgment about what might be distrusted in a jury's determination of pain-and-suffering awards—the possibility that isolated juries will diverge from population-wide convictions, the difficulty faced by lay people in generating a dollar number for certain classes of injuries, or something else.

## 2. Libel

Similar issues arise in the law of libel, which notoriously lacks clear measures of damages.<sup>205</sup> In fact the common law rules governing libel reflect the difficulty of generating monetary amounts.<sup>206</sup> Juries are asked to decide how much loss has been inflicted as a result of reputational injury. Thus, plaintiffs are able to recover both for identifiable pecuniary loss ("special damages") and also for damages stemming from general reputational harm that cannot be correlated easily with monetary measures ("general damages").<sup>207</sup> Sometimes plaintiffs are allowed to recover "presumed" general damages, which are awarded without proof of any actual damages, special or general.<sup>208</sup> Evidence from the plaintiff and the defendant is not likely to establish this amount with any accuracy. (If a movie star has been said to have engaged in adultery, what kind of award would provide actual compensation?)

---

204. See Kenneth S. Abraham & Glen O. Robinson, *Aggregative Valuation of Mass Tort Claims*, LAW & CONTEMP. PROBS., Autumn 1990, at 137, 141-46; Baldus et al., *supra* note 106, at 1123; Patricia M. Danzon, *Medical Malpractice Liability*, in LIABILITY: PERSPECTIVES AND POLICY 101, 121-24 (Robert E. Litan & Clifford Winston eds., 1988); Peter H. Schuck, *Scheduled Damages and Insurance Contracts for Future Services: A Comment on Blumstein, Bovbjerg, and Sloan*, 8 YALE J. ON REG 213, 215-19 (1991); cf. Geistfeld, *supra* note 199, at 828-43 (advocating the use of an ex ante full compensation award, including the possible use of damage schedules, for pain and suffering).

205. See Richard A. Epstein, *Was New York Times v. Sullivan Wrong?*, 53 U. CHI. L. REV. 782, 803-13 (1986).

206. See MARC A. FRANKLIN & DAVID A. ANDERSON, MASS MEDIA LAW 212-24 (5th ed. 1995).

207. See *id.* at 214.

208. See *id.*

In practice, the resulting verdicts are unlikely to draw a sharp line between compensatory and punitive damages.<sup>209</sup>

In this context, the jury is again likely to be mapping a complex judgment about the quality of the harm and perhaps the nature of the plaintiff and the defendant onto an unbounded dollar scale.<sup>210</sup> The results of combining a psychological understanding of juries with an evaluation of the goals of libel law will even bear on the law of free speech. It is not at all clear that the significant problem with libel law, for a system of free expression, consists of findings of liability, which might be accompanied by simple retractions; the most serious problem is probably the award of exorbitant sums of money.<sup>211</sup> An understanding of the sources of any such exorbitant awards will in turn bear on constitutional judgments about the relationship between libel and the First Amendment. What reforms would be appropriate?<sup>212</sup> Damage caps would have the same kind of crudeness for libel as in the context of punitive damages. Perhaps the three kinds of reforms discussed above, suitably adapted for the purposes of libel law, could be explored to reduce the unpredictable quality of libel judgments, or to make them a steadier and more realistic reflection of the actual goals, compensatory and deterrent, of the law of libel.

### 3. *Intentional Infliction of Emotional Distress and Sexual Harassment*

The latter half of the twentieth century has witnessed the rise of two important new legal wrongs: intentional infliction of emotional distress<sup>213</sup> and sexual harassment.<sup>214</sup> Both of these torts are accompanied by damage remedies. With respect to such remedies, the basic story should be familiar. Monetization is extremely difficult. Significant arbitrariness is entirely to be expected. Similar cases may well give rise to dramatically different awards. How does a jury know what amount would provide an employee, or a student, with adequate compensation for quid pro quo or hostile environment harassment? In both of these contexts, compensatory and punitive damages are

---

209. See Anderson & MacCoun, *supra* note 152 (presenting a study showing leakage between punitive and compensatory awards).

210. There is also a predictable difference between the amount that would compensate a libel plaintiff for injury inflicted and the amount that would persuade a libel plaintiff to allow his reputation to be damaged in the relevant way. See McCaffery et al., *supra* note 31 (examining how framing jury instructions in accordance with the two views may affect awards).

211. See FRANKLIN & ANDERSON, *supra* note 206, at 337-90.

212. See David A. Anderson, *Reputation, Compensation and Proof*, 25 WM. & MARY L. REV. 747 (1984) (proposing that alleviating harm to reputation should be the only function of defamation law); William W. Van Alstyne, *First Amendment Limits on Recovery from the Press—An Extended Comment on "The Anderson Solution,"* 25 WM. & MARY L. REV. 793 (1984) (proposing moving away from tort analysis in assessing libel claims); cf. Stanley Ingber, *Rethinking Intangible Injuries: A Focus on Remedy*, 73 CAL. L. REV. 772 (1985) (proposing a modified damage scheme limiting recovery for intangible injuries to the value of their pecuniary effects).

213. For an overview, see W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 12, at 54-66 (5th ed. 1984).

214. See generally CATHARINE A. MACKINNON, *SEXUAL HARASSMENT OF WORKING WOMEN* (1979).



likely to be entangled, in the sense that juries probably do not sharply separate the one from the other.

With both intentional infliction of emotional distress and sexual harassment, there may well be a relatively uniform set of underlying judgments among different demographic groups. For instance, it would be interesting to see whether there are differences between men and women or among other groups. This is an intriguing and entirely feasible empirical project along the lines of our study here. A principal source of unpredictability is likely to involve the translation of the underlying moral judgments into dollar amounts. Here, too, reform strategies might be based on a particular conclusion about what is wrong with the outcomes of jury deliberations—unpredictable awards, inadequate understanding of the effects of dollar amounts, or a reliance on improper factors.

#### 4. *Compensatory Versus Punitive Damages: General Considerations*

We can bring together some of the strands of this discussion by noting how the reform proposals discussed above may or may not bear on compensatory damage awards that are especially likely to be erratic. The most important feature of compensatory damages is that they are intended to restore the status quo ante.<sup>215</sup> Punitive damages, by contrast, are intended to reflect a normative judgment about the outrageousness of the defendant's conduct (together with a judgment about deterrence). Thus, the compensatory decision, far more than the punitive decision, reflects an assessment of fact (at least in theory).<sup>216</sup> At first glance, this is a sharp distinction between the two. In this light, would it make sense to consider reforms designed respectively to (1) capture a population-wide judgment about appropriate compensation (predictable populism); (2) capture a "compensatory intent" that would be mapped, by experts, onto dollar amounts (technocratic populism); and (3) dispense partly or entirely with juries on the ground that juries are unlikely to have the competence to make accurate judgments about the factual questions involved (bureaucratic rationality)?

To sort out these issues it is necessary to ask why juries are now charged with the task of making judgments about appropriate compensation in cases in

---

215. There is an important difference, though, between the amount a plaintiff would require to deem herself restored, and the amount a plaintiff would demand to incur the injury in the first instance. See *supra* note 203.

216. There are many complications here, some of them addressed below. An obvious issue is what, in this context, "compensation" is compensation for. If someone has suffered a month of pain, is compensation supposed to restore the plaintiff hedonically? Is it supposed to give dollar equivalents for injury to capabilities and functionings, to be assessed in part objectively rather than subjectively? See AMARTYA SEN, *COMMODITIES AND CAPABILITIES* (1985), for a defense of a "capability" approach to an assessment of well-being. Because the idea of "compensation" does not answer such a question, the jury's assessment inescapably creates normative issues.

which that inquiry strains their capacities. The most straightforward answer is self-consciously populist. In cases involving libel, pain and suffering, sexual harassment, and the intentional infliction of emotional distress, no institution is likely to be especially good at uncovering the “fact” about compensation, if there is indeed any such “fact.” Moreover, it is appropriate (on this view) to let the underlying decision reflect not merely facts but also the judgments of value held by the community as a whole.<sup>217</sup> Whatever factfinding deficiencies the jury may have (as compared to, say, a specialized agency) are overcome by the value of incorporating community sentiments into the decision about appropriate compensation for injuries that are not easily monetized. On this view, compensatory judgments are not so different from punitive judgments after all—both of them have important normative components.

Thus, the simplest argument on behalf of jury judgments about compensation is that any such judgment is—perhaps inevitably and certainly appropriately—not solely compensatory. It has evaluative dimensions, both in deciding what compensation properly includes and in imposing burdens of proof and persuasion and resolving reasonable doubts. The evaluative judgments, it might be thought, should be made by an institution with populist features and virtues.

The point may well apply to judgments about compensation for pain and suffering, libel, intentional infliction of emotional distress, and sexual harassment. A populist institution, on this view, should be permitted to undertake evaluative judgments about what amount would “compensate” someone who has suffered as a result of an improper medical procedure, a lie about his private life, or an unwanted sexual imposition by an employer or teacher. In the relevant cases, however, the problem of erratic judgments emerging from magnitude scaling without a modulus remains. This problem would not be severe (indeed, it would not be a problem at all) if what appeared to be erratic judgments were really a product of careful encounters with the particulars of individual cases, producing disparate outcomes that are defensible as such because they are normatively laden. But our study offers grave reasons to doubt that this is in fact the case.<sup>218</sup> Thus reformers might consider how the proposals discussed above would work in this context. As a

---

217. An underlying question in all of these areas is the extent to which the damage judgment should be person-specific. Suppose, for example, that an especially sensitive plaintiff has suffered an especially severe hedonic loss as a result of libel or sexual harassment—or, by contrast, that an especially tough-skinned plaintiff has suffered an unusually small hedonic loss as a result of the same torts. Should a jury consider the extent to which the plaintiff’s injury was objectively reasonable, independent of purely hedonic factors? Officially, tort law incorporates a reasonable person inquiry at the level of liability, but having been found liable, the defendant must take the plaintiff as the plaintiff experienced the injury. See *KEETON ET AL.*, *supra* note 213, § 65, at 453-54 (discussing contributory negligence). In other words, damages determinations are supposed to be person-specific. But we do not know if juries are willing to think in these terms, and it is unclear even that they should.

218. See also Sunstein, *supra* note 150.

starting point, we should notice that for compensatory damages, ranking is far preferable to rating along a bounded scale; it is certainly useful to see how a jury believes that the injury at issue compares with other injuries, and rating along a bounded scale is far less useful when punishment is not involved. If the basic problem is erratic judgments in the context of compensatory damages, it might be desirable to use a conversion formula to obtain a population-wide judgment about appropriate compensation.

A problem with this approach is that a population-wide judgment about appropriate dollar compensation might be ill-informed; it might not reflect "true" compensation. If the normative dimensions of that judgment seem to deserve a good deal of weight—if we see the jury's judgment about compensation as appropriately reflecting considerations not involving the apparently factual question of "compensation"—then this approach might well make sense. But if the factual dimensions deserve to predominate, the jury's ranking might be understood as a kind of "compensatory intent," to be converted to compensatory awards, not by population-wide data, but instead by an administrative or legislative conversion formula. Such a formula would be rooted in a judgment about the appropriate treatment of the cases against which the case at hand has been ranked. This kind of reform seems somewhat awkward, for the notion of "compensatory intent," which is supposedly rooted in a judgment about the facts, is less straightforward than that of "punitive intent," which is an unmistakably normative judgment. But it would mix populist and technocratic elements in a way that is mildly reminiscent of the treatment of social security disability cases, though in those cases administrative law judges displace juries.<sup>219</sup>

If the social security disability cases are truly a good analogy, technocratic considerations should predominate, and the third reform proposal would be best. On this view, an administrative or legislative body might create a kind of "pain-and-suffering grid," "libel grid," or "sexual harassment grid," combining the basic elements of disparate cases into presumptively appropriate awards.<sup>220</sup> A judge would produce a dollar award by seeing where the case at hand fits in the grid and perhaps by making adjustments if the details of the case strongly call for them. A technocratic approach of this kind could eliminate or at least greatly reduce the problem of erratic awards.<sup>221</sup> Whether

---

219. See *supra* text accompanying note 188. Note in this regard that many administrative agencies impose civil and criminal penalties, and they are also in a position to scale without a modulus. It would be extremely valuable to study their practice and to discover whether they have created some of the same kind of variability as that discussed here. See Rubin, *supra* note 42 (manuscript at 2-14) (discussing the similarity between punitive damages and administrative penalties).

220. Compare the analogous proposal for additur and remittitur in Baldus et al., *supra* note 106.

221. There is also an underlying question about the relationship between rule-bound judgment and particularistic judgment. Standards laid down in advance may leave room for erratic particularistic judgments if they are open-ended, but if they are rigid and rule-like they may prevent the reasonable exercise of discretion to adapt to the particulars of the individual case. The concern is how to minimize both decision costs and error costs; in the abstract, it is hard to know how much constraint on particularistic

it is desirable depends on the value of incorporating populist elements in the way that the more modest reforms would do.

Elements of these various approaches can be found in reform proposals, thus far restricted to the pain-and-suffering context, that attempt to cabin the jury's judgment by requiring it to decide in accordance with damage schedules and to place the case at hand in the context of other cases.<sup>222</sup> In view of the fact that similar problems beset other areas of law, there is no reason not to consider similar reforms in the contexts of libel, sexual harassment, and intentional infliction of emotional distress. A key issue is the appropriate role of technocratic and populist elements in the compensatory judgment. A judgment about that issue will go a long way toward shaping reforms.

## B. *Other Valuation Problems*

### 1. *Regulatory Expenditures*

In the last decade, there has been a great deal of interest in the problem of setting priorities for regulatory expenditures, both public and private. The pollutant-of-the-month syndrome has given rise to a fear that priorities are set in a random fashion and hence that expenditures per life saved are unpredictable.<sup>223</sup> Disparities between different live-saving programs are quite common and very substantial, to the point where reallocation of resources could save 60,000 lives per year (given the same investment as is currently made) or \$31 billion per year (given the same number of lives as are currently saved).<sup>224</sup>

Frequently, then, government must decide how much to expend per unit of regulatory benefit, and the phenomena that we have discussed here will likely come into play. Arbitrariness results partly from the difficulty of mapping normative judgments onto dollar amounts, and many regulatory problems are assessed in a no-comparison condition. Some of the most prominent efforts at regulatory reform can be understood as self-conscious responses. The attempt to value regulatory benefits through the "willingness-to-pay" criterion, for instance, introduces a budget constraint, with accompanying comparative judgments, into governmental decisionmaking.<sup>225</sup> And Justice Breyer's influential proposal of a special regulatory working group, entrusted with the task of allocating resources to large problems rather than small ones,

---

judgment will accomplish that task. For a good discussion, see Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992).

222. See, e.g., Bovbjerg et al., *supra* note 164, at 953; Saks et al., *supra* note 140, at 246.

223. See, e.g., STEPHEN BREYER, *BREAKING THE VICIOUS CIRCLE* 3-15 (1993).

224. See Tammy O. Tengs et al., *Five-Hundred Life-Saving Interventions and Their Cost-Effectiveness*, 15 RISK ANALYSIS 369 (1995); see also W. KIP VISCUSI, *FATAL TRADEOFFS* 261-63 (1992) (discussing differences in expenditure per life saved).

225. See VISCUSI, *supra* note 224, at 42-49.

is of a piece with his concerns about rule-free punitive damage awards.<sup>226</sup> His proposal is designed to ensure risk comparisons and to allow dollar judgments to be made by people with a good deal of experience in the task of “mapping” normative judgments onto a dollar scale. Justice Breyer’s proposal is in this sense parallel to the third kind of reform strategy that we have traced.

Justice Breyer’s proposal has been criticized as excessively technocratic.<sup>227</sup> Perhaps risk regulation has multiple goals, and the extension of human life does not adequately capture them.<sup>228</sup> Drawing on the second reform strategy described above—one that attempts to elicit the community’s normative judgments—we can imagine initiatives that would ensure a greater role for population-wide normative judgments while also promoting more expert “mapping” onto dollars.<sup>229</sup>

## 2. *Prison Terms*

Our emphasis throughout has been on an unbounded dollar scale, but the legal system makes use of another scale: years. Criminal punishment requires a decision about how to map a normative judgment onto a scale of years. That scale is not unbounded, in the sense that capital punishment (or life imprisonment) may be taken as an extreme end; but it presents a similar difficulty of scaling without an obvious modulus.<sup>230</sup> Before the enactment of the Sentencing Guidelines, there were serious problems of arbitrary and unpredictable sentences,<sup>231</sup> leading to dissimilar treatment of the similarly situated. It is reasonable to think that some of these problems resulted from the difficulty of mapping normative judgments onto a scale of years.

Is it true that, with respect to criminal punishment, people have predictable judgments about outrageousness and intention to punish, but unpredictable judgments about sentencing? Our findings here suggest the possibility of an affirmative answer, but much work remains to be done on this question. The answer obviously bears on the need for, and appropriate content of, any sentencing guidelines, and also on the general question of sentencing reform. Those who challenge the Sentencing Guidelines might be taken to be

---

226. See BREYER, *supra* note 223, at 59-72.

227. See, e.g., Lisa Heinzerling, *Political Science*, 62 U. CHI. L. REV. 449 (1995) (reviewing BREYER, *supra* note 223).

228. See *id.* at 470-71; Richard H. Pildes & Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. CHI. L. REV. 1, 86-88 (1995).

229. This is a goal of Pildes & Sunstein, *supra* note 228, at 86-95.

230. See STEVENS, *supra* note 137, at 31-33 (using magnitude scaling on jail sentences and using cross-modality matching against the severity of crimes).

231. See SANDRA SHANE-DUBOW ET AL., U.S. DEP’T OF JUSTICE, SENTENCING REFORM IN THE UNITED STATES (1985); Jon O. Newman, *Federal Sentencing Guidelines: A Risk Worth Taking*, BROOKINGS REV., Summer 1987, at 29.

complaining, among other things, about the absence of an appropriate modulus around which to organize diverse sentences.<sup>232</sup>

### 3. *Contingent Valuation*

The topic of contingent valuation, of course, raises the question whether people can turn their judgments about regulatory goods into nonarbitrary dollar awards.<sup>233</sup> This problem is closely related to the issue of punitive damages in the sense that it involves the mapping of a kind of attitude, desire, or judgment onto an unbounded dollar scale.<sup>234</sup> The goal of contingent valuation methods is to decide how much to value goods that are not traded on markets. Some commentators think that such valuation might be ascertained by exploring how people value goods that are in fact traded on markets.<sup>235</sup> But judgments about how much to spend to reduce statistical risks are highly contextual, and it is not clear that a decision to purchase a smoke alarm tells us a sufficient amount about how much people are willing to pay to reduce (for example) a risk of death from excessive levels of sulfur dioxide. The use of contingent valuation methods is inspired by a desire to obtain more specific, contextual assessments. Rather than looking at actual choices, these methods ask people hypothetical questions about how much they would be willing to pay to avoid certain harms or conditions.<sup>236</sup>

Despite their apparent promise, contingent valuation methods have serious limitations arising from the difficulty of mapping normative judgments onto dollars and the existence of framing effects. A special problem is that of indifference to quantity, or inadequate sensitivity to scope, reflected in the fact that people sometimes give the same dollar number to save 2000, 20,000, and 200,000 birds<sup>237</sup>—or the same number to save one, two, or three wilderness areas.<sup>238</sup> Consider the fact that Toronto residents are willing to pay almost as much to maintain fishing by cleaning up the lakes in a small area of Ontario

---

232. See, e.g., Albert W. Alschuler, *The Failure of Sentencing Guidelines: A Plea for Less Aggregation*, 58 U. CHI. L. REV. 901 (1991) (arguing that it is difficult to quantify situational and offender characteristics).

233. See the various perspectives in Symposium, *Contingent Valuation*, J. ECON. PERSP., Fall 1994, at 3.

234. There is also a question whether it is sensible to assume that people have well-formed judgments on such questions.

235. See VISCUSI, *supra* note 224, at 17-23.

236. See WILLIAM H. DESVOUGES ET AL., MEASURING NONUSE DAMAGES USING CONTINGENT VALUATION: AN EXPERIMENTAL EVALUATION OF ACCURACY 1-15 (Research Triangle Monograph No. 92-1, 1992); Robert Fabian et al., *Design of Contingent Valuation Approaches to Serious Illness*, in VALUING HEALTH FOR POLICY 278, 289-94 (George Tolley et al. eds., 1994); Symposium, *supra* note 233.

237. See DESVOUGES ET AL., *supra* note 236, at 71-73.

238. See Peter A. Diamond & Jerry A. Hausman, *Contingent Valuation: Is Some Number Better Than No Number?*, J. ECON. PERSP., Fall 1994, at 45, 51-52; Daniel Kahneman & Ilana Ritov, *Determinants of Stated Willingness To Pay for Public Goods: A Study in the Headline Method*, 9 J. RISK & UNCERTAINTY 5, 18 (1994).

as they are willing to pay to maintain fishing in all Ontario lakes.<sup>239</sup> Similarly, the valuation of a resource is affected by whether it is offered alone or with other goods. Willingness to pay for spotted owls drops significantly when the spotted owl is asked to be valued with and in comparison to other species. It is pertinent in this connection that the order and number of questions seem crucial in determining valuation. When asked for their willingness to pay to preserve visibility in the Grand Canyon, people offer a number three times higher when this is the first question than when it is the third question.<sup>240</sup>

What unifies contingent valuation and punitive damage assessment is the problem of mapping a normative judgment onto an unbounded dollar scale. There are illuminating contrasts between the two areas. Contingent valuation studies are understood as a method of ascertaining private willingness to pay for public goods.<sup>241</sup> By contrast, punitive damages are awarded on the basis of a community judgment of some sort; jurors are not asked their willingness to pay for any commodity. They are asked to generate a dollar amount that reflects, not the value of what has been lost, but that provides adequate deterrence or reflects social opprobrium about bad conduct. On the other hand, in both cases similar biases and distorting influences may be at work. As noted, both contexts present problems of mapping morally laden valuations onto dollar amounts. People may have predictable and nonrandom judgments about which species are most important, for example, and they may be able to compare various bodies of water, but a judgment about monetary valuation may be essentially arbitrary. And in both areas, the acontextual character of isolated judgments should make individual judgments less reliable.<sup>242</sup> In fact, our findings here have a parallel in the key conclusions of a study of contingent valuation, which also found that rating scales (e.g., of importance of the problem, moral satisfaction with contributing to its solution, or support for government action) are highly correlated with willingness to pay, and confirmed that the amount of systematic variance was much higher for willingness to pay than for these

---

239. See Daniel Kahneman & Jack L. Knetsch, *Valuing Public Goods: The Purchase of Moral Satisfaction*, 22 J. ENVTL. ECON. & MGMT. 57, 58-59 (1992).

240. See Note, "Ask a Silly Question . . .": *Contingent Valuation of Natural Resource Damages*, 105 HARV. L. REV. 1981, 1987 n.43 (1992) (citing a report to the EPA, GEORGE S. TOLLEY & A. RANDALL, ESTABLISHING AND VALUING THE EFFECTS OF IMPROVED VISIBILITY IN THE EASTERN UNITED STATES (1983)).

241. Of course, a large question is whether the answers received in the contingent valuation setting do reflect private willingness to pay in the market sense or reflect something else.

242. Judgments are also sensitive to framing effects. When asked how much they are willing to pay to prevent a species from being lost, subjects will offer a much lower amount than when asked how much they would have to be paid to allow a species to be lost. See RICHARD THALER, *QUASI-RATIONAL ECONOMICS* 167-77 (1991).

scales.<sup>243</sup> Furthermore, transformations to logs or ranks improve things there,<sup>244</sup> just as they do here.<sup>245</sup>

As with punitive damages, it may well make sense to consider substitutes for the current system of contingent valuation, perhaps rooted in the same considerations that we have discussed here.<sup>246</sup> Perhaps policymakers could develop a small number of scenarios for environmental damages, or use public judgments on a bounded scale, to begin a process by which such judgments might be translated into dollar amounts. New issues could be valued by a survey of attitudes that would include "explicit comparisons to the scenarios of the original scale."<sup>247</sup> Rather than being asked for dollar values, people would make assessments in terms of importance, and those judgments would then be mapped onto dollars by reference to the standard scale. Obviously, proposals of this kind raise complex questions that we cannot resolve here; we signal the issue only to emphasize the linkage with punitive damages and the general problem of mapping in the face of (potentially) shared moral judgments and (likely) erratic dollar amounts.

## VI. CONCLUSION

Why are jury determinations about punitive damages sometimes erratic and arbitrary? A large part of the answer lies in the difficulty of mapping normative judgments, including those of outrage and punishment, onto dollar amounts. This answer operates against an important backdrop: With respect to judgments of both outrage and punishment, important domains of law show substantial agreement in normative judgments, and the consensus operates across differences of gender, race, age, education, and income.

The fact that people have difficulty making judgments on an unbounded scale of dollars helps fortify and specify the basis for Justice Breyer's complaint, grounded in rule-of-law considerations, that some jurisdictions do not provide sufficient constraints on jury discretion.<sup>248</sup> The point also suggests that the solution does not lie in clearer jury instructions. Any effort to increase predictability in the award of punitive damages will be successful if and only if it assists with the task of mapping. If community judgments matter, legal reform should attempt to elicit the community's punitive intent and do so in terms of some response mode other than dollars.

When the legal system translates punitive intent into dollars, it must answer questions about the extent to which the law should incorporate, qualify,

---

243. See Kahneman & Ritov, *supra* note 238, at 14, 16.

244. See *id.* at 13.

245. See Kahneman et al., *supra* note 1 (manuscript at 38 tbl.6).

246. See Kahneman & Ritov, *supra* note 238, at 29-30; Rutherford et al., *supra* note 183, at 63-72.

247. Kahneman & Ritov, *supra* note 238, at 30.

248. See *BMW of N. Am., Inc. v. Gore*, 116 S. Ct. 1589, 1604-09 (1996) (Breyer, J., concurring).



or work against the jury's determination. We have suggested various reforms that embody different answers to those questions. Any ultimate conclusion depends on a specific assessment of what is wrong with current punitive damage awards. We have suggested three general possibilities: sheer variability, inadequate assessment by ordinary people of what different dollar awards will accomplish, or a focus by ordinary people on improper factors as the foundation for punitive awards. Predictable populism, technocratic populism, civil sentencing, and bureaucratic rationality are possible responses. Thus, an assessment of the normative issues requires an identification of the nature of the populist ideals that underlie the institution of the jury and a judgment about what place, exactly, those ideals deserve to have in light of juror psychology.

Our study shows that the characteristics of jury judgments include high sensitivity to outrage (and likely low sensitivity to the probability of detection), substantial sensitivity to harm, substantial sensitivity to firm size, susceptibility to anchors, and a backward-looking focus on retribution. A translation phase might incorporate or reject one or all of these characteristics; it may or may not be founded on the jury's punitive intention. At a minimum, our study strongly suggests that appellate judges and district courts should continue the practice, found in some courts,<sup>249</sup> of rejecting punitive awards that are out of line with general practice and relevant comparison cases—especially because, as we have shown, the dollar awards of any particular jury are not likely to reflect the population's judgment about appropriate dollar awards. A more dramatic approach—very possibly the best for the long term, though not without risks of its own—would involve the development of a system of administrative penalties for serious misconduct, based on judgments made in advance and subject to democratic control.

Our findings have implications well beyond the area of punitive damages. Three implications are of special importance. First, there is, in the personal injury cases studied here, a high degree of agreement within diverse demographic groups with respect to both outrage and punishment. It is possible that there is a similar consensus across other domains; thus ample room remains for further related work. It would be valuable to see, for example, whether such a consensus exists across gender with respect to damage awards for pain and suffering or sexual harassment. It would also be valuable to see whether there are substantive areas in which race, education, age, and wealth have measurable effects. Relatively simple studies could produce evidence on some much-disputed questions.

Second, our study raises serious doubts about whether jurors are able to make sensible judgments about dollar awards in cases in which market

---

249. See cases cited *supra* note 105. As noted, our evidence suggests that if community-wide judgments are the goal, it is important to have a mechanism for additur as well as remittitur.

measures are unavailable, and it throws into sharp relief the question of whether this task should be given to judges or to some other institution. The legal system is not occasionally but pervasively in the business of requiring people to map their judgments onto dollar amounts, and outside of the context of punitive damages, the translation is likely to suffer from the same problems found here. In particular, difficulties with mapping probably affect jury awards of damages for libel, pain and suffering, sexual harassment and other civil rights violations, and intentional infliction of emotional distress—all of which notoriously suffer from unpredictability, in part, we believe, as a result of unbounded dollar scales. The consequence is that rule-of-law values are badly compromised. Reform proposals might ensure against unfairness (in the form of dissimilar treatment of the similarly situated) and unpredictability for both individuals and firms, plaintiffs and defendants alike. More particularly, reforms involving “compensatory” awards in these controversial areas must be undertaken with an assessment of the appropriate place of purely factual judgments, for which juries are not especially well-suited. The three kinds of punitive damage reforms may well have parallels in each of these areas.

Third, and more broadly, our findings support the general proposition, on which there is growing consensus, that both values and preferences are often constructed, rather than elicited, by social situations.<sup>250</sup> In their various social roles, people lack a preexisting “master list” of values and preferences from which to make selections in situations of choice.<sup>251</sup> This is true for consumer choices, which depend on context, anchors, and alternative options, as well as for medical and political decisions. It is true as well for judgments made by jurors, judges, representatives, and citizens generally.<sup>252</sup> It is true, in particular, for judgments about punitive damages, which depend on response mode, available anchors, and the existence and nature of comparison cases. Rather than emerging from some menu in the mind, human judgments are a function of procedure (including response mode), description, and context. In the legal context, in particular, participants are pervasively in the business of constructing procedures, descriptions, and contexts. Frequently they have not been self-conscious about that point.

To be sure, there is no uncontroversial way to develop the context and frames for eliciting, or constructing, social preferences and values. But some ways are worse than others, because they make people perform tasks for which

---

250. See, e.g., Amos Tversky, *Rational Theory and Constructive Choice*, in *THE RATIONAL FOUNDATIONS OF ECONOMIC BEHAVIOUR* 185, 195 (Kenneth J. Arrow et al. eds., 1996); John Payne, J.R. Bettman & David Schkade, *Measuring Preferences in a Constructive World: Toward a Building Code* 3-9 (Nov. 4, 1997) (unpublished manuscript, on file with the *Yale Law Journal*).

251. See Anthony Bastardi & Eldar Shafir, *On the Pursuit and Misuse of Useless Information*, 75 *J. PERSONALITY & SOC. PSYCHOL.* (forthcoming July 1998) (finding that choices are constructed differently once people have started to pursue information).

252. See Kelman et al., *supra* note 31, at 289-303 (providing evidence of context-dependence in legal decisionmaking).

they are ill-equipped. More specifically, they produce poor translation of some plausible “bedrock” set of normative judgments, generate arbitrariness, aggravate the problem of selective attention, exploit heuristic devices that produce error, or provide people with too little understanding to yield sensible or consistent decisions. We have suggested some ways to handle these problems with the award of punitive damages, where bedrock judgments do seem discoverable through the right methods, and where the legal system allows those judgments to be transformed into dollar amounts in erratic ways. Both the problem and the potential solutions bear on many issues now facing the legal system.

## APPENDIX A: EXPERIMENTAL DESIGN

A. *Sample and Procedure*

The sample included diverse representation of various demographic and socioeconomic groups. For example, respondents were 44% male, and were 64% Caucasian, 16% African-American, and 15% Hispanic. The median income was between \$30,000 and \$50,000, and the median educational level attained was "some college." The median age of participants was between thirty and thirty-nine. Thirty-two respondents were dropped from the sample because they gave incomplete responses or failed to understand the task.

The survey was conducted at a downtown hotel and was administered to large groups at prearranged times over a four-day period. Most respondents completed their task in thirty to forty-five minutes.

Each respondent received three pages of general instructions and four numbered envelopes. The first three envelopes contained the materials for Parts 1, 2, and 3 of the study, as described below. The fourth envelope contained demographic questions and debriefing information. The instructions (excerpted in Appendix B) included (1) an overview of the survey procedure; (2) an explanation of the task of jurors in civil (as opposed to criminal) trials; (3) definitions of, and distinctions between, compensatory and punitive damages, including the fact that compensatory damages had already been awarded in the cases they would consider; (4) a summary of standard legal conditions for punitive damages (maliciousness or reckless disregard for the welfare of others); and (5) a reminder about the standard of evidence required in this situation (preponderance of the evidence).

B. *Design and Stimuli*

Ten scenarios describing personal injury cases were constructed. The first six were used in Parts 1 and 2 of the procedure and the other four in Part 3. Each respondent rated some version of all ten scenarios. Envelope 1 contained material about one of the scenarios. Envelope 2 contained five other scenarios. Envelope 3 contained the four scenarios used in Part 3 of the experiment.

All scenarios viewed by a given subject shared the same firm size and response mode. These factors were varied across respondents, as was the order in which the scenarios appeared. Scenario order was counterbalanced using a Latin square so that each scenario appeared in each ordinal position with equal frequency. Table 6 shows the wording of the evaluation questions in the three response modes. Instructions in all scenarios stated that compensatory damages of \$200,000 had already been awarded.

Part 3 had the same structure as Parts 1 and 2, except that the isolation manipulation was replaced by a manipulation in which the degree of harm suffered by the plaintiff was varied. For each of the four scenarios used in Part 3, we formulated a high-harm and a low-harm version. For example, in the case in which a child playing with matches was burned when his pajamas caught fire, the injuries were described as follows: "He was severely burned over a significant portion of his body and required several weeks in the hospital and months of physical therapy" (high harm), or "His hands and arms were badly burned, and required regular professional medical treatment for several weeks" (low harm).

## APPENDIX B: EXCERPTS FROM INSTRUCTIONS

In this study we would like you to imagine that you are a juror for a legal case in a civil court. Civil lawsuits can involve disputes between private individuals, companies, or individuals and companies, in which the plaintiff alleges that the defendant harmed them or their property in some way. A civil suit is brought by a plaintiff for the purpose of gaining compensation from the defendant for the alleged harm.

Civil suits involve two different types of penalties that could be imposed upon a defendant that is found liable for damages. Compensatory damages are intended to compensate fully a plaintiff for the harm suffered as a result of the defendant's actions. Punitive damages are intended to achieve two purposes: (1) to punish the defendant for unusual misconduct; and (2) to deter the defendant and others from committing the same actions in the future.

In the cases you will consider, the defendant has already been ordered to pay compensatory damages to the plaintiff. This does not necessarily mean that punitive damages must also be awarded. Whether or not punitive damages should be awarded and, if so, how large they should be, is completely separate from compensatory damages.

Punitive damages should be awarded if a preponderance of the evidence shows that the defendants either acted maliciously or with reckless disregard for the welfare of others. Defendants are considered to have acted maliciously if they intended to injure or harm someone. Defendants are considered to have acted with reckless disregard for the welfare of others if they were aware of the probable harm to others but disregarded it, and their actions were a gross deviation from the standard of care that a normal person would use.

Civil suits differ from criminal cases, in which the government prosecutes an individual or a company for alleged violations of the law. Plaintiffs in a civil trial must prove their claim by "a preponderance of the evidence," which means that it is more likely than not that the plaintiff's claim is justified. This differs from criminal trials, where the prosecution must prove the defendant's guilt "beyond a reasonable doubt."

In each of the cases you will consider, the jury (of which you are a member) has already decided to accept the plaintiff's claim. As a consequence the jury has ordered the defendant to pay \$200,000 in compensatory damages to the plaintiff as full compensation. The defendants are large (medium-sized) companies with profits of \$100-200 [\$10-20] million per year.

## APPENDIX C: PERSONAL INJURY SCENARIOS

*Mary Lawson*

Mary Lawson, a manufacturing worker, developed chronic anemia. Although after a hospital stay she is now better, the condition has not been fully cured. She believes that exposure to benzene in her workplace caused the condition and sued her employer, TGI International. The jury (of which you are a member) ordered TGI International to pay her \$200,000 in compensation.

TGI International is a large company (with profits of \$100-200 million per year) that manufactures high-tech machine parts. Some years ago, the scientists at TGI International discovered that manufacturing workers at Mary Lawson's plant were often exposed to benzene, a substance that can

cause anemia, leukemia and cancer. Internal documents show that the top management at TGI International decided not to do anything about the problem, because benzene levels in the plant were slightly below the maximum level allowed by OSHA regulations. They thought that the risk was worth taking and that "with any luck no one will get sick." They also decided against warning the workers, because "warnings would just create panic."

*Frank Williams*

Frank Williams suffered serious internal injuries when the braking system on his motorcycle failed to work at a traffic light. He felt that the brakes were defective, and sued National Motors, the company that manufactures and sells his motorcycle. The jury (of which you are a member) ordered National Motors to pay him \$200,000 in compensation.

National Motors is a large company (with profits of \$100-200 million per year) that makes motorcycles, scooters, and other motorized single-person vehicles. The evidence showed that the braking system used by National Motors has a basic design defect. National Motors was aware that "there might be a problem with our brakes," because in premarket tests, the defect appeared on several occasions. But the premarket tests were not extensive, despite the fact that auto industry regulations require elaborate testing. Internal company documents show National Motors's belief that "it would be quite expensive for us to do much more now, we can't be certain we have a serious problem here, and anyway we can fix the problem afterwards if it really does turn out to be serious."

*Thomas Smith*

While he was visiting the circus, Thomas Smith was shot in the arm by a security guard who mistakenly thought that Smith had threatened another customer with bodily harm. The security guard was drunk at the time. Smith sued Public Entertainment, the company that operates the circus. The jury (of which you are a member) ordered Public Entertainment to pay him \$200,000 in compensation.

Public Entertainment is a large company (with profits of \$100-200 million per year) which operates circuses and public fairs. Fred Williams, the security guard who was involved in the incident, is an alcoholic with a history of incidents of drunkenness on the job. During one of these incidents, Williams took out his gun and started waving it around wildly, but he did not shoot anyone. Public Entertainment had repeatedly warned Williams to "clean up his act" but took no other action. In his company personnel file, Williams was described as "basically a good guy with a bit of a drinking problem, but not enough of a risk to fire him."

*Susan Douglas*

Susan Douglas suffered significant injuries to her legs and neck when an airbag in her car opened unexpectedly while she was driving the vehicle. She believes that the airbag was defective and sued the manufacturer, Coastal Industries. The jury (of which you are a member) ordered Coastal Industries to pay her \$200,000 in compensation.

Coastal Industries is a large company (with profits of \$100-200 million per year) that specializes in parts and accessories that can be added to existing vehicles, such as adding the latest safety equipment

to older cars. While its airbag conforms to the requirements stated in government regulations, it does not include certain additional "fail-safe" systems that are used in other airbags to ensure against accidental opening. Internal documents show that most but not all of the Coastal Industries designers believed that their system "is certainly safe enough, even if it does not include all possible safeguards" and that their marketing department said that "there will be no market for our airbag if we raise its price by adding more safety bells and whistles."

#### *Carl Sanders*

Carl Sanders used Nalene, an over-the-counter baldness treatment available at drugstores. While a small amount of hair did grow back, he also developed severe side-effects, including open sores on the scalp and permanent brown spots over his forehead. He sued the manufacturer, A&G Cosmetics. The jury (of which you are a member) ordered A&G Cosmetics to pay him \$200,000 in compensation.

A&G Cosmetics is a large company (with profits of \$100-200 million per year) that sells many different cosmetic products, including wigs, "weaves," and chemical solutions designed to combat baldness. Nalene has proven effective in promoting hair growth in 30% of people in clinical trials. However, Nalene caused unpleasant side effects in some cases, although none was as severe as those Carl Sanders experienced. When marketing Nalene, A&G Cosmetics did not fully disclose these findings. It only said "minor side effects have been observed in a very small number of people tested." While this amount of disclosure was within legal limits, other companies that make hair products voluntarily disclosed more about their products.

#### *Sarah Stanley*

Sarah Stanley, a seventy-five year old woman, suffered serious back injuries as a result of following an exercise video, "Good Health For All," that she purchased through her local community health center. When Stanley attempted to perform the exercises, she found herself unable to do so, but she pressed on beyond her physical capacities. She claimed that she was not adequately warned of these dangers and sued the producer of the video, Gersten Productions. The jury (of which you are a member) ordered Gersten Productions to pay her \$200,000 in compensation.

Gersten Productions is a large company (with profits of \$100-200 million per year) that produces informational materials in health-related fields, including videos on many topics concerning healthy lifestyles. The "Good Health for All" video contains a series of exercises suitable mostly for people in good shape and good health. The exercise coaches and models in the video are all relatively young, and no federal or state law requires exercise videos to come with any special warning for elderly people. The witnesses in the case testified that Gersten Productions believed that most people would be able to tell when the exercises were beyond their capacities, that "Good Health for All" has produced good results for almost all people who have seen it, and that very few people had reported injuries of any type from doing so.

*Jack Newton*

Jack Newton, a five-year-old child, was playing with matches when his cotton flannel pajamas caught fire. He was severely burned over a significant portion of his body and required several weeks in the hospital and months of physical therapy. His parents sued the manufacturer of the pajamas, Novel Clothing. The jury (of which you are a member) ordered Novel Clothing to pay the Newtons \$200,000 in compensation.

Novel Clothing is a large company (profits of \$100-200 million per year) that specializes in making clothes for children. Before marketing the pajamas, Novel conducted the tests normally used in the industry for problems like flammability, and observed no incidents like the Newtons experienced. Companies in the industry as well as federal regulators have known for a while that it is possible to add extra fire-retardant chemicals to their fabrics (in addition to those specified in current regulations), but these extra measures are not required. The process is very expensive, and no other manufacturers currently use it. Internal documents show that the management of Novel Clothing had decided that "when it comes to costly safety innovations we will follow our competitors. We don't want to be less safe than anyone else but we don't have to lead the way either."

*Low-Harm Version:* Jack Newton, a five-year-old child, was playing with matches when his cotton flannel pajamas caught fire. His hands and arms were badly burned, and he required regular professional medical treatment for several weeks.

*Joan Glover*

Joan Glover, a six-year-old child, ingested a large number of pills of Allerfree, a nonprescription allergy medicine, and required an extensive hospital stay. The overdose weakened her respiratory system, which will make her more susceptible to breathing-related diseases such as asthma and emphysema for the rest of her life. The Allerfree bottle used an inadequately designed child-proof safety cap. The Glovers sued the manufacturer of Allerfree, the General Assistance company. The jury (of which you are a member) ordered General Assistance to pay the Glovers \$200,000 in compensation.

General Assistance is a large company (with profits of \$100-200 million per year) that manufactures a variety of nonprescription medicines. A federal regulation requires "child-proof" safety caps on all medicine bottles. General Assistance has systematically ignored the intent of this regulation by selling tens of thousands of bottles of medicines with a child-proof safety cap that was generally effective, but had a failure rate much higher than any other in the industry. An internal company document says that "this stupid, unnecessary federal regulation is a waste of our money"; it acknowledges the risk that Allerfree may be punished for violating the regulation but says that "the federal government has many other things to worry about and probably won't bother us on this" and in any case that "the punishments for violating the regulation are extremely mild; basically we'd be asked to improve the safety caps in the future." An official at the Food and Drug Administration had previously warned a vice president of General Assistance that they were "on shaky ground" but the company decided not to take any corrective action.



*Low-Harm Version:* Joan Glover, a six-year-old child, ingested a large number of pills of Allerfree, a nonprescription allergy medicine. She had to spend several days in a hospital, and is now deeply traumatized by pills of any kind. When her parents try to get her to take even beneficial medications such as vitamins, aspirin, or cold remedies, she cries uncontrollably and says that she is afraid.

*Martin West*

Martin West, a right-handed disabled veteran who lived in a two-story house, was seriously injured in a fall when the chain broke on his electric lift-chair (a device that allows someone to be carried up stairs in a chair that moves up and down an angled track). He fell from near the top of the stairs and tumbled awkwardly all the way to the bottom landing, damaging his spinal cord in the process. As a result, he now has only partial control of his right arm, a condition which doctors believe is permanent. He sued the manufacturer of the lift-chair, MedTech Products. The jury (of which you are a member) ordered MedTech to pay him \$200,000 in compensation.

MedTech Products is a large company (profits of \$100-200 million per year) that manufactures many types of medical equipment, including wheelchairs, car-lifts, and other devices used by the disabled. The lift-chair is a new product for MedTech, and instead of producing a new design, company engineers decided to adapt the design of the hydraulic lifts for cars already on the market. Unfortunately, there are several unique problems in designing a safe and effective lift-chair that were beyond the experience of the company's engineers. Product managers said that hiring new engineers with the proper expertise was "too expensive, and would take too long" and ordered current engineers to "just do the best you can, but be sure you meet our deadline for announcing the product." The inexperience of the engineers and the rush to meet the product announcement date led to testing procedures that were less rigorous than those required by federal medical product regulations.

*Low-Harm Version:* Martin West, a left-handed disabled veteran who lived in a two-story house, was injured in a fall when the chain broke on his electric lift-chair (a device that allows someone to be carried up stairs in a chair that moves up and down an angled track). He fell from near the bottom of the stairs and tumbled to the bottom landing, injuring his spinal cord in the process. His right arm was paralyzed for several weeks, after which doctors were able to repair most of the injury, and he was able to regain most of the previous range of motion in the arm.

*Janet Windsor*

Janet Windsor, a secretary who works on computer equipment, developed a rare form of skin cancer. After a long course of painful chemotherapy, doctors were able to cure the cancer, although they cannot be sure that it will not return. She believed that it had been caused by the computer monitors that she worked on and sued the manufacturer, International Computers. The jury (of which you are a member) ordered International Computers to pay her \$200,000 in compensation.

International Computers is a large company (profits of \$100-200 million per year) that manufactures components of computer systems. The type of International Computers monitor that Ms. Windsor used emits an unusually high level of radiation compared to other similar monitors, a level that pushes the limit in government safety guidelines. Internal company documents cite experts who

concluded that “the evidence that this level of radiation could create any serious risk to health and life is weak and tentative.” The company was not legally required to disclose the unusual level of radiation, and it did not do so.

*Low-Harm Version:* Janet Windsor, a secretary who works on computer equipment, suffered from frequent and severe migraine headaches. As a result, for several years she often experienced nausea, insomnia and depression, and missed many workdays and family events.

#### APPENDIX D: TABLES

TABLE 5. SUMMARY OF PERSONAL INJURY SCENARIOS

CASE	PART	DESCRIPTION
Mary	1, 2	Employee suffers anemia due to benzene exposure on the job
Frank	1, 2	Motorcycle driver injured when brakes fail
Thomas	1, 2	Circus patron shot in arm by drunk security guard
Susan	1, 2	Auto airbag unexpectedly opens, injuring driver
Carl	1, 2	Man suffers skin damage from using baldness cure
Sarah	1, 2	Elderly woman suffers back injuries from using exercise video
Jack	3	Small child playing with matches burned when pajamas catch fire
Joan	3	Child ingests large quantity of allergy medicine, needs hospital stay
Martin	3	Disabled man injured when wheelchair lift malfunctions
Janet	3	Secretary chronically ill due to radiation from computer monitor

TABLE 6. RESPONSE MODE MANIPULATION

#### Outrage

Which of the following best expresses your opinion of the defendant's actions? Please circle your answer.

Completely				Absolutely		
Acceptable		Objectionable		Shocking		Outrageous
0	1	2	3	4	5	6

*Punishment*

In addition to paying compensatory damages, how much should the defendant be punished? Please circle the number that best expresses your opinion of the appropriate level of punishment.

No Punishment	Mild Punishment	Severe Punishment	Extremely Severe Punishment			
0	1	2	3	4	5	6

*\$ Damages*

In addition to paying compensatory damages, what amount of punitive damages (if any) should the defendant be required to pay as punishment and to deter the defendant and others from similar actions in the future? Please write your answer in the blank below.

\$ \_\_\_\_\_

