

1997

Assessing Punitive Damages...

Cass R. Sunstein

Daniel Kahneman

David Schkade

Follow this and additional works at: https://chicagounbound.uchicago.edu/law_and_economics



Part of the [Law Commons](#)

Recommended Citation

Cass R. Sunstein, Daniel Kahneman & David Schkade, "Assessing Punitive Damages..." (Coase-Sandor Institute for Law & Economics Working Paper No. 50, 1997).

This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

ASSESSING PUNITIVE DAMAGES
(WITH NOTES ON COGNITION AND VALUATION IN LAW)

Cass R. Sunstein,¹ Daniel Kahneman,² and David Schkade³

I. INTRODUCTION

A. Two Goals

This Article has two goals. The first is to investigate the sources of arbitrariness in punitive damage awards.⁴ This investigation, which we believe to be the first of its kind, helps to answer an important question: What are the sources of unpredictability in punitive damage awards?⁵ On the basis of

¹ Karl N. Llewellyn Distinguished Service Professor, Law School and Department of Political Science, University of Chicago.

² Eugene Higgins Professor of Psychology and Professor of Public Affairs, Princeton University.

³ Professor of Management and William W. Spriegel Fellow, Graduate School of Business, University of Texas, Austin. The authors are grateful to Exxon Corporation for financial support of the empirical study contained in this Article; Exxon Corporation bears no responsibility for our analysis or our conclusions. We are grateful to Richard Craswell, Richard Epstein, Dan Kahan, William Landes, Tracey Meares, Martha Nussbaum, Eric Posner, and Richard Posner for valuable comments. We are also grateful to participants in the rationality workshop at the University of California, Berkeley and the law and economics and legal theory workshops at the University of Chicago. J.B. Heaton provided research assistance and excellent comments.

⁴ A technical version of the survey and study, accompanied with more detailed statistical analysis, is reported in Daniel Kahneman, David Schkade, and Cass R. Sunstein, *Shared Outrage and Erratic Awards: The Psychology of Punitive Damages*, *J. of Risk and Uncertainty* (forthcoming 1998). This paper provides a nontechnical summary of our findings and develops, as that paper does not, its implications for law and legal reform.

⁵ We do not contend that we have identified the only source. See TAN *infra*. There are also questions about the difference between experimental and real-world conditions. See TAN *infra*.

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, are that *people's moral judgments are remarkably widely shared, but that people have a great deal of difficulty in mapping their moral judgments onto an unbounded scale of dollars*. 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 answer 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 libel, awards for pain and suffering, 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, 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 last three decades, a great deal of progress has come from the application, to legal problems, of a certain understanding of economics.⁶ Within economics and psychology, but outside law, that understanding has been under sustained attack.⁷ 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.⁸

⁶ See Richard Posner, *Economic Analysis of Law* (5th ed. 1997) for a wide range of relevant discussions.

⁷ A good recent collection is William Goldstein and Robin Hogarth, *Research on Judgment and Decisionmaking* (1997).

⁸ There are some important exceptions. See, e.g., Edward McCaffery, Matthew Spitzer, and Daniel Kahneman, *Framing the Jury*, 81 *Va. L Rev*

These insights very much bear on law. Our study helps show what might be missing, impractical, or wrong in standard economic approaches to punitive damage awards, and in a way that bears on a range of additional issues as well.⁹

B. Shared Outrage and Erratic Awards

The award of punitive damages has become one of the most controversial and important uses of the tort law, extending well beyond the common law to such statutory areas as environmental protection and employment discrimination.¹⁰ Punitive damages are allowed in forty-seven of the fifty states,¹¹ and over sixty federal statutes now permit the award of punitive damages, making judicial review of punitive awards a significant part of federal law.¹²

The purposes of such awards are not obscure. Sometimes compensatory awards provide insufficient deterrence of

1341 (1995); Mark Kelman, Yuval Rottenstreich, and Amos Tversky, Context-Dependence in Legal Decisionmaking, 25 J. Legal Stud. 287 (1996); George Loewenstein and Linda Babcock, Explaining Bargaining Impasse, 11 J. Econ. Persp. 109 (1997).

⁹ See A. Michell Polinsky and Steven Shavell, Punitive Damages: An Economic Analysis, Harv L Rev (1997) (forthcoming). In brief, the difficulty with the Polinsky/Shavell suggestion—that juries should decide punitive damage problems by reference to the likelihood that the defendant's acts will go undetected—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 they suggest. Thus a reader might conclude from the Polinsky/Shavell analysis that punitive damages should be assessed by an administrative agency, or some other expert body, and not by juries or even judges. See TAN *infra*.

¹⁰ See the Civil Rights Act of 1991, 42 U.S.C. 1977A.

¹¹ See Richard Blatt et al., Punitive Damages: A State by State Guide to Law and Practice (1991).

¹² See, e.g., 7 U.S.C. (a)(11)(3); 7 U.S.C. 18(a)B); 7 U.S.C. 21 (b) (10); 7 U.S.C. 25(a)(3)(B); 10 USC 2207 (a)(2); 10 USC 2409(c)(C)(2); 15 USC 1681n(2); 15 USC 1681u(i) 15 USC 2622((b)(2)(B); 18 USC 248(cc)(1)(B)); 18 USC 25209(b)(2); 18 USC 2724 (b)(2); 42 USC 300j-9; 42 USC 300a-23; 42 USC 1981a; 42 USC 3613; 42 USC 7622; 42 USC 9607; 42 USC 13981.

private behavior, simply because injured parties do not detect and seek compensation for all injuries. Punitive damages can, in theory, take account of the infrequency of private suits by penalizing the defendant to the point where it will undertake optimal precautions.¹³ Moreover, punitive damages may have a retributive or expressive function, designed to embody social outrage at the actions of serious wrongdoers.¹⁴ They may reflect the “sense of the community” about the egregious character of the defendant’s action.

Understood in these terms, however, punitive awards raise several puzzles. Such awards interact in complex ways with public law—with the elaborate network of regulatory requirements and criminal prohibitions that overlap with the tort law. These requirements and prohibitions 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.

In recent years, the most important concern is that whatever their purposes, punitive damages are unpredictable, even “out of control.”¹⁵ One study of 47 counties in eleven states over a several-year period showed a high degree of variability:¹⁶ Punitive damages were awarded in about 25% of the successful verdict cases in some counties and not awarded at all in others.¹⁷ Median verdicts ranged from less than \$10,000 in some areas to as much as \$204,000 in San

¹³ See William Landes & Richard Posner, *The Economic Structure of Tort Law* 160-65, 184-85, 223-24 (1993); Polinsky & Shavell, *supra* note.

¹⁴ See David Owen, *The Moral Foundations of Punitive Damages*, 40 *Ala L Rev* 705 (1989); Marc Galanter and David Luban, *Poetic Justice*, 42 *Am. U. L. Rev.* 1393 (1993); Jean Hampton, *The Retributive Idea*, in Jean Hampton and Jeffrie Murphy, *Forgiveness and Mercy* 111 (1988).

¹⁵ John Jeffries, *A Comment on the Constitutionality of Punitive Damages*, 72 *Va L Rev* 139, 139 (1986).

¹⁶ Stephen Daniels and Joane Martin, *Myth and Reality in Punitive Damages*, 75 *Minn L Rev* 1 (1990).

¹⁷ *Id.* at 32.

Diego.¹⁸ Another study showed increasingly large awards against businesses, from a total of \$1.1 million in five large states from 1968 to 1971 to \$343 million from 1988 to 1991, an 89-fold increase.¹⁹ A recent study finds that in San Francisco, California and Cook County, Illinois, punitive damage awards increased from about \$1 million during 1960-64, to \$147 million during 1985-1989, and \$215 million during 1990-94.²⁰

On the basis of evidence of this kind, many people have complained that punitive awards have a “lottery-like” character²¹ and require legislative or judicial remedy.²² Thus it is said, in well-publicized documents, that the arbitrary character of punitive damage awards is “news to no one,” that they produce an affront to the rule of law by “distributing awards in a random and capricious manner,”²³ and that they should be subject to a cap of three times the actual harm.²⁴

¹⁸ *Id.* at 42.

¹⁹ See Turner et al., *Punitive Damages Explosion*, Washington Legal Foundation, Working Paper Series No. 50 (Nov. 1992).

²⁰ Deborah Hensler and Erik Moller, *Trends in Punitive Damages: Preliminary Data from Cook County, Illinois and San Francisco, California* (RAND Institute for Civil Justice unrestricted draft series, March 1995). See also Rand Institute for Civil Justice, *Punitive Damages in Financial Injury Verdicts* (1997) (showing substantial variability and growth in punitive awards).

²¹ Peter Huber, *No-Fault Punishment*, 40 *Ala L Rev* 1037, 1037 (1989).

²² See Paul Rubin, John Calfee, and Mark Grady, *BMW v. Gore: Mitigating the Punitive Economics of Punitive Damages*, 1997 *Supreme Court Economic Review* 179, 184; E. Donald Elliott, *Why Punitive Damages Don't Deter Corporate Misconduct Effectively*, 40 *Ala. L. Rev.* 1053, 1057 (1989). See the attempted response in the Product Liability Reform Act of 1997, S. 5, 105th Cong., 1st Sess. (1997) (proposing to require a clear and convincing evidence standard and also to cap punitive damage awards).

²³ *Contract with America: The Bold Plan* by Rep. Newt Gingrich, Rep. Dick Arme, and The House Republicans to Change the Nation 154 (Ed Gillespie and Bob Schellhas eds., 1994).

²⁴ *Id.* at 147.

Congress continues to debate bills that would limit the availability and amount of punitive damages.²⁵

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.²⁶ Hence unpredictable awards create both unfairness and (on reasonable assumptions) inefficiency,²⁷ in a way that may create overdeterrence of desirable activity.²⁸

²⁵ See, e.g., H.R. 956, 104th Cong, section 2(a) (1995); H.R. 955, 104th Cong, section 8(a), (C) (1995).

²⁶ See Kenneth MacCrimmon and Donald Wehrung, *Taking Risks: The Management of Uncertainty* (1986).

²⁷ These concerns about punitive damage awards have not produced a consensus that punitive damages are unpredictable. Some empirical work suggests that the aggregate data show a reasonably orderly pattern, though 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). A response, coming to a different conclusion on the basis of the same data, is A. Mitchell Polinsky, *Are Punitive Damages Really Insignificant, Predictable, and Rational?*, 26 *J. Legal Stud.* 663 (1997).

In their careful analysis, Eisenberg et al. show that the logarithmic transformation yields a distribution that is almost normal. They also show that the log of awards is predicted rather well from a set of objective characteristics of cases in which such awards were made. Eisenberg et al. conclude that the unpredictability of punitive awards has been overstated. We find no inconsistency between their analyses of real jury awards and our experimental data. Indeed, we agree with their conclusion that log

Our principal interest here is in identifying some of the sources of unpredictability in jury judgments. On the basis of a study of 899 jury-eligible citizens, we offer the following major findings.

1. People have a remarkably high degree of moral consensus on the degree of outrage and punishment that are appropriate for punitive damage cases.²⁹ 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-female, all-Hispanic, all-male, all-poor, all-wealthy, all-black juries, all-old juries, 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

awards are fairly predictable. But defendants and plaintiffs live in a world of dollars, not of log dollars. In terms of 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 skewness creates the possibility of either modest or disastrous losses in identical cases. This produces unfairness, because the similarly situated are treated differently, and also might induce risk aversion even in very large firms.

²⁸ For relevant evidence see Paul H. Rubin et al., *BMW v. Gore: Mitigating the Punitive Economics of Punitive Damages*, 5 *Supreme Court Economic Review* 179, 192-96 (1997); Richard L. Manning, *Products Liability and Prescription Drug Prices in Canada and the United States*, 40 *J. L. & Economics* 203, 210-34 (1997).

²⁹ Two qualifications are necessary. First, this conclusion is restricted to the area we investigate, involving products liability 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. See below.

punitive intent are turned into highly erratic judgments about appropriate dollar punishment. A basic source of arbitrariness with 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 in what we call “a no-comparison condition.” 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. This effect is, however, far less important than the effect identified in (2) in producing arbitrary awards.

The principal purpose of this Article is to set out and 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.³⁰ Thus there is reason to ask whether the civil justice system ought not to be brought more closely in line with the criminal justice system, where juries of course decide questions of liability, and judges decide questions

³⁰ Of course there is a large question about the importance of predictability in an assessment of 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 they do not reflect the right theory 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. See TAN *infra*. But we discuss below the relationship between unpredictability and other possible problems with damage awards.

about punishment, subject to guidelines and constraints. If juries cannot consistently or sensibly “map” their judgments onto an unbounded dollar scale, might it follow that judges, rather than juries, should be making decisions about punitive awards? Should the legal system shift to a system of *civil sentencing*? Might the same conclusion make sense for compensatory awards in cases in which dollar awards are also likely to be arbitrary? Might the current system of civil liability, both punitive and compensatory, sometimes be displaced by an administrative process, one that can aspire to more in the way of coherence and rationality?

The choice of reform of course depends on an assessment of precisely what 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 case 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. 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 those of the population as a whole. The result would be a form of *predictable populism*.

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, rather than 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 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 problem 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, administrative penalties, and judgments about the appropriate focus of the regulatory state. About one-half of tort law involves monetary awards for injuries that are hard to monetize.³¹ From our findings, it is reasonable to infer that the phenomenon of shared judgments and erratic or arbitrary³² dollar amounts create a serious and unaddressed problem in many areas. Very typically, juries and judge are mapping judgments onto an unbounded dollar scale; the phenomenon of widespread and predictable judgments, combined with the demonstrable cognitive difficulty of translating preferences and values into dollar amounts, has wide-ranging implications for the operation of both private and public law.

Most generally, our findings raise a simple 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 some progress on it here.

This Article comes in six parts. Part II outlines existing understandings about punitive damages, showing how the sources of variable judgments are relevant to constitutional issues and also to existing theories of deterrence and retribution. Part III outlines our study and the basic conclusions. Part IV presents policy recommendations designed to provide a role for community judgments without the “noise” and arbitrariness that accompany the current system. What is important is not the details of the proposals

³¹ See W. Kip Viscusi, *Reforming Products Liability* 102-104 (1993).

³² They may be arbitrary either in the sense of erratic or in the sense of a product of a normatively questionable “anchor” on which the jury has seized. See TAN *infra*.

but their basic goal and direction: to provide juries with an opportunity to do what they are capable of doing relatively well, without requiring juries to do what they are bound to do badly.

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 call for an ambitious decision about 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. Part VI is a conclusion.

II. DETERRENCE, RETRIBUTION, AND THE CONSTITUTION

A. Policy Notes

It is not our purpose to resolve the debate about whether and why a court should award punitive damages.³³ Our modest goal is to outline some of the relevant arguments by way of introduction. It is important to provide those arguments here, since they provide the foundation for the constitutional law governing punitive damage awards, and because the arguments play a large role in our empirical study.

³³ Relevant discussions include Polinsky and Shavell, *supra* note; Dorsey Ellis, *Fairness and Efficiency in the Law of Punitive Damages*, 56 *Southern Cal. L. Rev.* 1 (1982); Robert Cooter, *Economic Analysis of Punitive Damages*, 56 *S. Cal. L. Rev.* 79 (1982); Robert Cooter, *Punitive Damages for Deterrence*, 40 *Alabama L. Rev.* 1143 (1989); David Owen, *The Moral Foundations of Punitive Damages*, 40 *Ala. L. Rev.* 705 (1989); Marc Galanter and David Luban, *Poetic Justice*, 42 *Am. L. Rev.* 1393 (1993).

The traditional view is that punitive awards serve deterrent and retributive goals. Thus 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."³⁴

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 and to bring suit against the injurer, and to 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, adequate deterrence will not be provided, since wrongdoers will be able to continue to engage in conduct whose social costs exceed social benefits. Punitive damages are necessary to pick up the slack for undetected wrongdoers.³⁵

Under this view, the goal of punitive damages is to ensure that the award of compensatory damages is supplemented by an amount reflecting the fact that the probability of that award is less than 100%. The simplest conclusion is that total damages should be the harm multiplied by the reciprocal of the probability than the

³⁴ Ronald Eades, *Jury Instructions of Damages in Tort Actions* 98 (3d ed. 1995).

³⁵ See William Landes and Richard A. Posner, *The Economic Structure of Tort Law* 160-65 (1987); Polinsky & Shavell, *Punitive Damages: An Economic Analysis*, *Harv L. Rev.* forthcoming.

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.³⁶ We will suggest some institutional and psychological reasons to doubt that jurors can or will make judgments of this kind; let us continue the simple account for now.

With some torts—medical malpractice is an obvious example—it is very plausibly the case that many defendants will be able to avoid 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 damages award in the *BMW* case (discussed below), because it is not easy to detect fraudulent repaintings 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 certain homicides and assaults.

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 has been, and should be, made that certain subjective gains ought not to be allowed to count at all.³⁷ If someone has gained utility from murder, rape, assault, or sexual harassment, it might be thought—reasonably enough—that that gain should not be permitted to count. Of course the argument that such gains should not count is not itself an economic judgment, but it is not hard to imagine defenses of that argument.³⁸ Punitive damages are necessary to offset the utility gain that is judged illicit. Compensatory damages are inadequate, in such cases,

³⁶ See Polinsky & Shavell, *Punitive Damages: An Economic Analysis*, *Harv. L. Rev.*, forthcoming.

³⁷ Some utilitarians make just this argument. See the discussion of the exclusion of sadistic or malicious preferences in John Harsanyi, *Morality and the Theory of Rational Behavior*, in *Utilitarianism and Beyond* 39 (Amartya Sen and Bernard Williams eds. 1982)

³⁸ See *id.*

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 reason for punitive damages stems from the fact that sometimes compensatory damages may be lower than they should be, and sometimes 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.³⁹ Of course this argument has a degree of speculativeness and also crudeness, because it is unlikely that punitive damages can be calibrated with sufficient precision to make up for the shortfall in compensatory damages. But perhaps it is reasonable to think that punitive damages bring the incentives of wrongdoers closer to where they should be.

On this view, intentional torts, involving deliberate infliction of an injury, may well provide good cases for punitive damages. Of course it is in such cases that socially illicit gains are most likely to be involved. In such cases, the probability of obtaining damages from the injurer may well be less than 100%; intentional torts frequently (though not always) involve a form of concealment. But an important implication is that punitive damages generally do not make sense for highly visible torts where the probability of detection and compensation is extremely high; there, compensatory damages are all that is required.⁴⁰

(b) *Puzzles and problems.* Even from the economic point of view, these arguments for punitive damages raise some serious puzzles. An important question, not yet taken up in the economic literature, is the relation between this rationale for punitive damages and the existence of other regulatory controls, for example those created by administrative and criminal law. Both administrative and criminal law are often

³⁹ See Landes and Posner, *supra* note, at 160.

⁴⁰ See *id.*; Polinsky and Shavell, *supra* note.

defended as an effort to compensate for private underenforcement of law;⁴¹ 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.

There is an independent point, closely related to our findings here. Even without administrative and criminal law, efforts to impose punitive damages may well misfire because of the inevitable confusion or cognitive and motivational errors of jurors and judges. There is thus a question whether real-world institutions should or can reliably engage in the enterprise of seeking to obtain optimal deterrence even if they seek to do so. 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.

A general problem is that 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.⁴² It is for this

⁴¹ See, e.g., Richard A. Posner, *Economic Analysis of Law* 220 (4th ed. 1994). A general discussion is Edward Rubin, *Punitive Damages: Reconceptualizing the Runcible Remedies of Common Law*, *Wisconsin Law Review* (forthcoming 1997).

⁴² Thus there is a possibility here of health-health tradeoffs, which occur when a strategy designed to reduce health risks actually increases health risks. Some punitive damage awards may actually have adverse health effects, if they serve to increase prices. For a recent discussion, see Richard L. Manning, *Products Liability and Prescription Drug Prices* in

reason that the strongest cases involve intentional torts, and 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 desirable activity by hypothesis is not 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.⁴³ 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.⁴⁴ The Supreme Court has said that wealth and income can be taken into account.⁴⁵ On a conventional view about optimal deterrence, however, wealth and income is 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

Canada and the United States, 40 J. L. & Economics 203, 210-34 (1997). For general discussion, see John Graham and Jonathan Weiner, *Risk vs. Risk* (1996); Cass R. Sunstein, *Health-Health Tradeoffs*, 63 U. Chi. L. Rev. 1533 (1996). Evidence of such effects, in the particular context of punitive damages, can be found in Paul Rubin, John Calfee, and Mark Grady, *BMW v. Gore: Mitigating the Punitive Economics of Punitive Damages*, 1997 Supreme Court Review 179, 192-96.

⁴³ For the view that wealth should be irrelevant, see *Kemezy v. Peters*, 79 F.3d 22 (7th Cir. 1996); Kenneth Abraham and John Jeffries, *Punitive Damages and the Rule of Law*, 18 J. Legal Stud. 415 (1989); Ellis, *supra* note, at 61-63; Polinsky and Shavell, *supra* note. For an argument the other way, see Jennifer Arlen, *Should Defendants' Wealth Matter?*, 21 J. Legal Stud. 413, 414 (1992) (discussing deterrence of risk-averse people).

⁴⁴ See TAN *supra*. Compare Robert MacCoun, *infra* note 45.

⁴⁵ See *infra*. Cf. Robert MacCoun, *Differential Treatment of Corporate Defendants by Juries*, 30 Law & Society Review 121, 133-39 (1996) (finding effects from corporate identity and commercial activity, but not from wealth *per se*).

engage in optimal behavior, whatever its wealth (assuming risk neutrality).⁴⁶ But there is a question whether this is realistic; we take up this issue below.

2. *Retribution*

Thus far we have spoken in terms of deterrence; but under conventional noneconomic analysis, 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,⁴⁷ 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 careful normative account of the relationship between retributive goals and punitive damages. Those who emphasize retribution are seeking to ensure that the community's outrage about certain acts is reflected in punishment. Probably the retributive idea would focus on two principal factors: the defendant's state of mind and the degree of harm actually caused or likely to be caused by defendant's behavior. The first point is especially important. Retributists who are moral egalitarians⁴⁸ attempt, through civil or criminal punishment, to capture the intuition that certain intentional states are especially bad because they reflect abhorrent and false views about the moral worth of persons.⁴⁹ In cases where one person murders another, or

⁴⁶ See Landes and Posner, *supra* note, at 163-64; Polinsky and Shavell, *supra* note.

⁴⁷ See TAN *infra*.

⁴⁸ Of course retributivists need not be egalitarians; we could imagine retributists who thought that harms against some persons deserve less punishment than harms against other persons.

⁴⁹ See Jean Hampton, *The Retributive Idea*, in Jean Hampton and Jeffrie Murphy, *Forgiveness and Mercy* 111 (1988).

acts with gross negligence that produces another's death, the wrongdoer reveals a belief that some people do not matter very much.⁵⁰ Public outrage is the appropriate response.

In this sense retribution, properly understood, embodies a principle of moral equality. The award of punitive damages can also be taken to have an expressive function; it expresses the community's outrage at certain kinds 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.⁵¹ With this understanding it is not surprising that punitive damages have come to be awarded for sexual harassment and for other violations of civil rights statutes;⁵² the relevant community now believes that these kinds of illegality reflect an especially bad state of mind, and hence that punitive damages are a necessary supplement to other sanctions. The defendant must be punished accordingly. In any event the retributive idea suggests that the most serious cases for punitive damage awards thus involve harms that are both grave in degree and affirmatively desired by the defendant, as, for example, in examples of homicide or assault.

We connect these factors to the psychology of punitive damage awards in the discussion below. The central point is that 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

⁵⁰ See *id.*; Jean Hampton, *Correcting Harms Versus Righting Wrongs: The Goal of Retribution*, 39 *UCLA L Rev* 1659 (1992). Hampton discusses punitive damage awards in particular in *id.* at 1687-89; see also the illuminating discussion in Marc Galanter and David Luban, *Poetic Justice: Punitive Damages and Legal Pluralism*, 42 *Am Univ L Rev* 1393 (1993) (connecting punitive awards with retributive goals and community outrage).

⁵¹ See Dorsey Ellis, *Fairness and Efficiency in the Law of Punitive Damages*, 56 *S Cal L Rev* 2, 14-15 (1982).

⁵² See, e.g., 42 U.S.C. 1977A.

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. Options

What is the relationship between the award of punitive damages and the Constitution? Because punitive damage awards have been so controversial, and because the Supreme Court has been effectively forced to evaluate a range of punitive damage judgments, this question has become one of the more complex and pressing in modern constitutional law. There are three possible answers: (a) Punitive damages awards are always constitutional. (b) Such awards are unconstitutional, as violations of the due process clause in its substantive dimension, when they are “grossly excessive.” (c) 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. Our findings here are directed principally toward proposition (c) and to the general question 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 high as 20%) as arbitrary because out of step with the facts of the case or with comparison cases.⁵³

A majority of the Supreme Court has recently converged on proposition (b), though propositions (a) and (c) have also received significant support, and though (c) is highly likely to be a prime area for future contestation.⁵⁴ In fact the majority’s

⁵³ See cases cited in notes *infra*.

⁵⁴ *BMW of North America v. Gore*, 116 S. Ct. 1589 (1996).

endorsement of a version of (b) is accompanied by an understanding that a major problem with excessive awards is that they fail to provide sufficient predictability.⁵⁵ Hence a majority of recent justices—most prominently Justices O'Connor and Breyer, but at different times also Chief Justice Rehnquist and Justices Souter, White, Brennan, Marshall, Stevens, Kennedy, and Blackmun—have argued that the due process clause requires constraints on jury discretion that will provide fair notice and limit the role of arbitrary or irrelevant factors. The Court has not given much of an account of how juries might be led in the direction of arbitrariness or unpredictability; 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 early cases the Court refused to set aside punitive damage awards as inconsistent with the excessive fines and due process clauses.⁵⁶ But it left open the possibility that in an extreme case, an award would be constitutionally unacceptable under the due process clause. Thus the punitive damages decisions have refused to endorse the rule, proposed most insistently by Justice Scalia,⁵⁷ that the Constitution imposes no constraints on what juries may do in this context. Instead the Court left open the possibility of a substantive due process limitation on excessive awards. The Court stressed the possibility of a successful claims on the basis of substantive rather than procedural due process insofar as the Court 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 capital punishment, where there is a similar judgment to be made

⁵⁵ *Id.* at 1604; *id.* at 1605-07 (Breyer, J., concurring).

⁵⁶ *TXO Producton v. Alliance*, 509 US 443 (1993); *Pacific Mutual v. Haslip*, 499 US 1 (1991).

⁵⁷ See *BMW of North America v. Gore*, 116 S. Ct. 1589 (Scalia, J., dissenting).

about whether the problem with the death penalty is that it is sometimes or always excessive or instead that it is produced by insufficiently reliable procedures. After seriously considering both routes, the Court eventually converged on the procedural option, and it is possible that this will happen with punitive damage awards as well. And as we will see, the choice of substantive rather than procedural due process is of great importance; empirical evidence suggests some advantages to the procedural route.

2. *Repainted cars*

For constitutional purposes, the key case is *BMW of North America v. Gore*.⁵⁸ In that case, Dr. Gore sought punitive damages because his new BMW had actually been repainted, and he was not informed of this fact. The jury granted an award of punitive damages of four million dollars, an amount that seemed well out of line with the \$4000 compensatory damages award. Presented by this disparity, the Court ruled for the first time that an award of punitive damages violates the due process clause. But there was an internal division. The opinion for the five-member majority spoke in terms of excessiveness and hence substantive due process,⁵⁹ though with subthemes of federalism and fair notice. Justice Breyer's concurring opinion was procedurally oriented.⁶⁰ The four dissenting justices seemed to say that no punitive damage award could ever violate the due process clause,⁶¹ part of the impetus for their opinion was to impose good incentives on democratic branches to take care of the problem.⁶² It is worthwhile to spend some time with the opinions, to show

⁵⁸ 116 S. Ct. 1589 (1996).

⁵⁹ *Id.* at 1598-1604.

⁶⁰ *Id.* at 1605-09.

⁶¹ *Id.* at 1610-1614 (Scalia, J., joined by Thomas, J., dissenting); *id.* at 1614-1618 (Ginsburg, J., joined by the Chief Justice, dissenting).

⁶² Compare the idea of a penalty default in the law of contracts and statutory construction. See note *supra*; see also Ian Ayres and Robert Gertner, *Filling Gaps in Incomplete Contracts*, 99 *Yale LJ* 87 (1989).

where our findings about jury judgments connect with existing law.

In finding the award grossly excessive, the Court began with the suggestion 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; and though it left some ambiguity on this point, it suggested that a state may impose punitive damages only in order to protect its own consumers and its own economy. 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."⁶³ 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":⁶⁴ the degree of reprehensibility of the nondisclosure; the relation between harm and potential harm on the one hand and the punitive damages award on the other; and the difference between the punitive award and available penalties for comparable misconduct. Taking these guideposts together, the Court found the award unconstitutional because "grossly excessive."

The first factor, in the Court's view, is an effort to ensure some proportionality, in which damages reflect the extent of the offense.⁶⁵ Here retributive goals appear central. In the Court's view, this was a serious problem with the jury's award, because no special aggravating considerations were present in Gore's case. The injury was purely economic. There was no effect on performance or safety of the car, or even its appearance for a significant period. There was no indifference to or reckless disregard for health and safety. The failure to

⁶³ 116 S. Ct. at 1598.

⁶⁴ *Id.*

⁶⁵ *Id.* at 1599.

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 outrageous behavior, deliberate false statements, concealment of evidence of improper motive, or affirmative misconduct.⁶⁶

In discussing the second factor, the Court said that the ratio of punitive damages to actual (or potential) compensatory damages was especially bad: over 500 to 1. This was worse than that in previous cases.⁶⁷ Importantly, the Court did not say that the ratio would be decisive. If an especially egregious act produced a small amount of economic damage, high punitive damages would be legitimate; so too where “the injury is hard to detect or the monetary value of noneconomic harm might have been difficult to determine.”⁶⁸ But 500 to 1 is “breathtaking.”⁶⁹

Third, 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 \$2000.⁷⁰ Thus the punitive damage award was quite inconsistent with legislative judgments about the relevant conduct as seen in other areas of the law. The point matters 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 principle of fair notice.⁷¹ For Justice Breyer, the most serious problem in the case was not sheer excessiveness but the absence of legal standards that could

⁶⁶ *Id.* at 1601.

⁶⁷ *Id.* at 1601. Compare the Haslip case, involving a ratio of over 4 to 1, *Pacific Mutual Insurance Co. v. Haslip*, 499 US 1 (1991), and also that in *TXO Production Corp. v. Alliance Resources Corp.*, 509 US 443 (1993), involving a ratio of 10 to 1.

⁶⁸ *Id.* at 1602.

⁶⁹ *Id.* at 1603.

⁷⁰ *Id.*

⁷¹ *Id.* at 1605.

reduce decisionmaker discretion and hence caprice. The central problem, for Justice Breyer, was unlimited jury discretion. Here the relevant standards “are vague and open-ended to the point where they risk arbitrary results.”⁷² The jury operated under no statute with standards distinguishing among permissible punitive damage awards. In the *Green Oil* case,⁷³ decided in 1989, Alabama set out seven factors to be used by appellate courts to constrain punitive damages awards. These included: a reasonable relation to actual and likely harm; degree of reprehensibility; removal of profit; the financial position of the defendant; costs of litigation and need to create incentives for private litigants; the imposition, or not, of criminal sanctions on the defendant; and the existence of mitigation from other civil penalties.⁷⁴ In this case, however, the seven factors were not applied in a way that made for actual constraint. Nor have the state courts made any effort to discipline those factors in such a way as to generate a legally constraining standard. In so saying, Justice Breyer referred to the economic test referred to above, in particular 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 lawsuits that would likely be brought, and adding generous attorneys fees and other costs.”⁷⁵ Here there was no evidence that the Alabama Supreme Court applied “any economic theory” to explain the \$2 million recovery.⁷⁶ Nor was there a community understanding or historic practice that would provide background standards exemplified in that recovery. The general problem lay in the violation of the Rule

⁷² Id. at 1608.

⁷³ *Green Oil Co. v. Hornsby*, 539 So.2d 218 (Ala. 1989).

⁷⁴ Id. at 223.

⁷⁵ 116 S. Ct. at 1607.

⁷⁶ Id.

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.”⁷⁷

Thus Justice Breyer’s opinion can be understood as connecting the outcome in *BMW* with void for vagueness cases⁷⁸ and the constitutional attack on the death penalty in *Furman v. Georgia*.⁷⁹ The central problem lies in unconstrained discretion. For Justice Breyer, the outcome in the *BMW* case is not best understood simply by reference to excessiveness. We will refer to Justice Breyer’s approach in a number of places below. The problems identified here connect directly with his procedural concerns; but they also suggest a distinctive source of discretion, one very different from that emphasized in his opinion.⁸⁰

⁷⁷ *Id.* at 1609.

⁷⁸ See, e.g., *Papachristou v. City of Jacksonville*, 405 US 156 (1972).

⁷⁹ 408 US 238 (1972).

⁸⁰ Two other cases should be briefly discussed. In *Pacific Mutual Life Insurance Co. v. Haslip*, 499 US 1 (1991), Haslip complained of the lapsing of his health insurance policy as a result of misconduct by one Ruffin, an agent for Pacific Mutual and also for another, unaffiliated company. Ruffin had misappropriated premiums issued by Haslip’s employer for payment to the other insurer. The trial court instructed the jury that it could award punitive damages against Pacific Mutual if it found liability for fraud. 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. The Court held that the award was acceptable: the instructions referred to deterrent and redistributive goals, pointed to the character and degree of the wrong, and excluded evidence of Pacific Mutual’s wealth; the Supreme Court of Alabama had established post-trial procedures for scrutinizing punitive damage awards; and that Supreme Court provided an additional check by examining whether there is a reasonable relation between the award the actual and likely harm, the degree and duration of the defendant’s conduct, the defendant’s state of mind, any concealment by the defendant, and the frequency of past similar conduct; the profitability to the defendant; and other factors. *Id.* at 17-20.

The Court was badly divided over the damage award in *TXO Production Corp. v. Alliance Resources Corp.*, 509 U.S. 443 (1993). *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

C. Valuation, Variability, and Constraints

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.⁸¹ 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, for example, that the likely 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 bad. And in *Honda Motor Co. v. Oberg*,⁸² the Court held, in a way that bears a good deal on our study, that the due process clause requires judicial review of the size of punitive damage awards, to ensure that they are not arbitrary or excessive. The Court said that a state that failed to provide judicial review violated

damages. The Court upheld the award. Justice Stevens, writing as well for Chief Justice Rehnquist and Justice Blackmun, wrote the plurality opinion, holding that the TXO award was not excessive merely because the potential harm was much higher than the actual harm. *Id.* at 459-62. 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. *Id.* at 462. In any case TXO's pattern of behavior threatened millions of dollars in losses to others. *Id.* Thus there was no grotesque disparity between punitive damages and threatened harm. 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. *Id.* at 474.

⁸¹ See, e.g., *Kim v. Nash Finch Co.*, 1997 US App LEXIS 22511 (1997); *Mathie v. Fries*, 1997 US App LEXIS 19874 (2d Cir. 1997).

⁸² 512 U.S. 415 ((1994).

procedural due process, because it risked a “lawless, biased, or arbitrary” result.⁸³

But all this leaves 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. But the *BMW* case practically forces lower courts to begin to do exactly that, and this is what is now emerging.⁸⁴

What the Court lacks, and what bears on constitutional controls, is an understanding of the source of jury variability. Where jurors’ judgments range over a wide range, what is the reason? Are jurors reacting to fine-grained judgments about 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. What are the standards for judicial review of punitive awards? Appellate courts, admittedly with constitutional pressure in the background, are often in the business of deciding whether to reduce punitive damage awards as unreasonable.⁸⁵ In fact a significant percentage of punitive awards do not survive appellate review.⁸⁶ To decide when to reduce awards,

⁸³ *Id.* at 433.

⁸⁴ *Id.*

⁸⁵ *Klein v. Grynberg*, 44 F. 3d 1497 (10th Cir. 1995); *Stafford v. Puro*, 63 F.3d 1436 (7th Cir. 1995); *Allahar v. Zahora*, 59 F.3d 693 (7th Cir. 1995); *Ross v. Black and Decker, Inc.*, 977 F.2d 1178 (7th Cir. 1992); *Vasbinder v. Scott*, 976 F.2d 118 (2nd Cir. 1992); *Kimzey v. Wal-Mart Stores, Inc.*, 107 F.3d 568 (8th Cir. 1997); *Lee v. Edwards*, 101 F.3d 805 (2nd Cir. 1996); *King v. Macri*, 993 F.2d 294 (2nd Cir. 1993); *Michelson v. Hamada*, 36 Cal.Rptr.2d 343 (Cal. Ct. App. 1994); *Wollersheim v. Church of Scientology*, 6 Cal.Rptr.2d 532 (Cal. Ct. App. 1992); *Baume v. 212 E. 10 N.Y. Bar Ltd.*, 634 N.Y.S. 2d 478 (N.Y. App. Div. 1995); *Parkin v. Cornell University Inc.*, 581 N.Y.S.2d 914 (N.Y. App. Div. 1992):

⁸⁶ See David Baldus et al., *Improving Judicial Oversight of Jury Damages Assessments: A Proposal for the Comparative Additur/Remittur*

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. It is predictable that 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 award⁸⁷—a question to which we will return. If variable and erratic judgments are the problem, the legal system should correct awards that are both unreasonably low and unreasonably high, for both of these are likely to occur.⁸⁸

Of course the whole question of punitive damages is under active consideration in many states as well as at the national level.⁸⁹ States are considering caps in the form of ratios or flat

Review of Awards For Nonpecuniary Harms and Punitive Damages, 80 Iowa L. Rev. 1109, 1120 n. 20 (1995).

⁸⁷ There is in some states a procedure for “additur,” see David Baldus et al., Improving Judicial Oversight of Jury Damages Assessments: A Proposal for the Comparative Additur/Remittur Review of Awards For Nonpecuniary Harms and Punitive Damages, 80 Iowa L. Rev. 1109, 1119-1120 (1995), 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*, 293 U.S. 474 (1934).

⁸⁸ The best defense of the current system—allowing 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, though it provides particular evidence of “skewing” in the form of high rather than low awards.

⁸⁹ Three states do not allow punitive damages: Nebraska, New

dollar limitations; many states are attempting to require bifurcated trials in which judges rather than juries determine the level of punitive damages.⁹⁰ These are somewhat crude and categorical efforts to respond to the general fear that punitive damages are “out of control.”⁹¹ An understanding of the source of variability would lead in the direction of 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: (i) the psychology of the sequence of judgments and

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* 18-26 (1988). Federal proposals include a cap on punitive damages in some civil suits, see S. 5, 105th Cong, 1st Sess. (1997). H.R. 56, 104th Cong section 201 (1995).

⁹⁰ State law reforms fall in several categories. See generally *Developments in the Law—The Civil Jury*, 110 Harv L Rev 1408, 1527-1536 (1997), for an overview. (1) Some states have enacted caps; other states are considering that same approach. Some caps limit the entire dollar amount; others limit punitive damage awards to a multiple of compensatory damages; others relate possible punitive amounts to the income of the defendant. (2) Some states requires part or all of the punitive award to go to state agencies or the state treasury. See, e.g., Fla. Stat. ch. 768.73/(2) (1995); Kan. Civ. Proc. Code Ann. 60-3402(e). (3) Connecticut, Kansas, and Ohio have required the judge, not the jury, to determine punitive damage awards. See Conn Gen State 52-240b; Kan Civ Proc Code 60-3701(a); Ohio Rev Code Ann 2315.21. (4) Some states have required a bifurcated trial, in a way that is designed to reduce confusion and to make clear which factors are relevant to which proceeding. (5) Over half the states have increased the standards of proof, from preponderance of the evidence to clear and convincing evidence or more.

⁹¹ 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. See below for more discussion.

attitudes that produce individual judgments about punitive awards in particular cases; (ii) the sources of variability in these judgments; (iii) the implications of these findings for the unpredictability of jury awards, and for possible reforms in the tasks assigned to juries.

A total of 899 respondents were recruited from the voter registration rolls of Travis County, Texas, and paid \$35 to participate in the study. A set of 10 vignettes of personal injury cases were created in which a plaintiff (always an individual customer) sued a firm for compensatory and punitive damages. All vignettes had 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 28 different variations of the 10 vignettes. Each respondent evaluated 10 cases, composed of one variation of each of the 10 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 sub-samples were asked to answer different questions about each scenario: how outrageous was the defendant's behavior (on a scale of 0 to 6); how much should the defendant be punished (on a scale of 0 to 6); how much should the defendant be required to pay in punitive damages. Each respondent first dealt with one case without seeing the others (the "no comparison" condition), then received a booklet with nine new cases.

A. The Outrage Model

We propose a descriptive theory of the psychology of punitive awards, called the outrage model.⁹² The essential claim is that the moral transgressions of others evoke an attitude of outrage, which combines an emotional evaluation

⁹² For further discussion see Kahneman, Schkade, and Sunstein, *supra* note.

and a response tendency. The rules that govern outrage present an important problem that we do not address in this paper. We assume that outrage is largely governed 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 the degree outrageousness on a numerical scale. Under the outrage model, punitive 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 sentences for attempted murder. The relationship between victim and juror was manipulated in an experiment by Hastie, Schkade, and Payne.⁹³ Under experimental conditions, larger awards were made when the plaintiff was located in the jurors' community than when the plaintiff was from a remote location. We speculate that this factor affects punitive intent: the retributive urge is stronger

⁹³ See Reid Hastie, David Schkade, and John Payne (forthcoming).

when the victim belongs to one's group than when the victim is a stranger.⁹⁴

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 just one of a number of possible scales on which outrage might be expressed. 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 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.

Our two central hypotheses were simple:

⁹⁴ The size of the awards was also affected by the amount demanded by plaintiff. 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.

1. Shared outrage

The outrage evoked by different scenarios of tortious behavior is governed by a bedrock of broadly shared social norms.

With respect to outrage and punitive intent, are randomly selected juries likely to be similar to one other? And are rankings of different scenarios generally similar for different demographic groups? We hypothesized affirmative answers to both questions, at least in the context of the products liability cases given here.

2. Erratic dollar amounts

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.

This prediction was tested 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:

3. The harm effect

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 was varied.

4. Firm size effect

Damage awards are determined by punitive intent and by other ascertainable factors, prominently including the size of the defendant firm.

This prediction was tested by presenting each scenario in two versions, in which the size of the defendant was varied.

5. Less discrimination without comparison

In a “no-comparison” condition, 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.

This hypothesis was tested by giving respondents cases in isolation and in the context of other cases.

B. 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, white, Hispanics, African-Americans, 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,⁹⁵ over the set of 28 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, African-Americans, and whites, and respondents at very different levels of income and education produced almost identical orderings of the 28 scenarios used in the study.⁹⁶ Judgments of intent to punish in these

⁹⁵ We used means for outrage and punishment and medians for dollar awards.

⁹⁶ Of course, considerable variability between individuals remains within a given demographic category, even though the aggregate responses are very similar to the aggregate responses of another

scenarios of product liability cases evidently rest on a bedrock of moral intuitions that are broadly shared in society. We also looked for differences among groups in the average severity of their judgments (i.e., the level of the average rating), and here too found no significant differences (with the one exception of gender, to be taken up below).

This striking finding may not generalize to all domains of the law.⁹⁷ We might expect to find larger differences between communities and social categories in other domains of the law, perhaps including attitudes toward civil rights violations and environmental harms; at least it is possible that, for example, African-Americans would *rate* civil rights cases more severely than whites do. We expect, however, that within the same category of cases, *rankings* may 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.

Table 1. Correlation Between Demographic Groups on Intended Severity of Punishment*

		<u>Men</u>	
Gender	Women	.974	
		<u>Black</u>	<u>White</u>
Ethnicity	White	.975	

category. Our analysis simply shows that identifiable groups have similar distributions, not that all individuals are alike.

⁹⁷ There is also a question whether the competing narratives of the real world of juries 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.

ASSESSING PUNITIVE DAMAGES

	Hispanic	.963	.988		
			<u>< 30K</u>	<u>30-50K</u>	
Household	30-50K	.991			
Income	> 50K	.986			
			<u>< 30</u>	<u>30-39</u>	<u>40-49</u>
Age	30-39	.994			
	40-49	.992		.994	
	> 50	.991		.993	.987

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

C. The Exception of Gender Ratings

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 plaintiff's behavior as more outrageous (a mean difference of .52 scale units, $p < .001$), expressed more punitive intent (a mean difference of .37 scale units, $p < .001$), and set higher log dollar awards ($p < .01$). 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 ($p < .05$).

This finding should not be overemphasized; the differences in ratings were relatively small. But it bears 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 fact that women tend to favor stiffer punishments—and that men tend to favor more lenient punishments—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 men do.⁹⁸

D. Unpredictable Dollar Awards

Our central hypothesis was that dollar 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

⁹⁸ See Paul Slovic, University of Chicago Legal Forum (forthcoming 1997).

replacement, groups of 12 responses to each case for each response scale. In this manner we 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,⁹⁹ and indeed may even understate our ultimate conclusion.¹⁰⁰

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

⁹⁹ See James Davis, *Group Decision Making and Quantitative Judgments*, in *Understanding Group Behavior* (E. Witte and James Davis eds. 1996); Shari Diamond and Jonathan Casper, *Blindfolding the Jury to Verdict Consequences*, 26 *Law and Society Review* 513 (1992). Clearly, the appropriateness of this measure may depend on the task structure of the group (e.g., whether or not a unanimous decision is required). However, in our study, replacing the median of juror’s individual judgments as the group decision with the mean, has little effect on the results for outrage and punishment, but makes dollar awards even less predictable.

¹⁰⁰ See Diamond and Casper, *supra* note; Davis, experiment 1, *supra* note; Martin Kaplan and Charles Miller, *Group Decision Making and Normative Versus Informational Influence*, 53 *J. Personality and Social Psychology* 306 (1987). Note, however, the existence of an “amplification of bias,” by which a group process, involving a set of individuals biased on one direction or another, may push awards in extreme directions, in fact more extreme than that of any individual before deliberation begins. We are indebted to Robert MacCoun for suggesting this possibility. Cf. N. Kerr, Robert MacCoun, and G. Kramer, *Bias in Judgment: Comparing Individuals and Groups*, 103 *Psych. Review* 687 (1996) (finding an amplification of bias, but not in the setting of punitive damage determinations). The possibility of extremes resulting from group deliberation would fortify our conclusion, by showing even greater variance.

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 between 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 are we to get the same ratings and rankings as in this first set of trials?”

To answer this question we conducted an analysis requiring 4 steps. (1) We created a randomly selected jury for each of the 28 cases, and computed the median judgment for each. (2) We then 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 28 randomly selected juries and corresponding median judgments. The correlation between these and the first set of jury judgments is a measure of how erratic or consistent juries are. (3) To get a more reliable indication of the typical correlation between juries, we performed Step 1 60 times for each of the three response modes. This produced 180 columns of data, each of which contained one set of 28 jury judgments. (4) We then computed the correlations between every pair of sets of simulated jury judgments (i.e., correlations between the columns). This computation was performed both within response (e.g., the correlation between two sets of 28 outrage ratings) and across responses (e.g., the correlation between a set of 28 outrage ratings and a set of 28 punishment ratings). The data shown in Table 3 are medians of the correlations

obtained for each response mode or response mode combination.

As we had hypothesized, the individual differences in dollar awards produce severe unpredictability, and highly erratic outcomes, even in the medians of 12 judgments (the results would be even more extreme with smaller samples such as 6-person juries). While there is strong agreement between independent sets of outrage or punishment juries ($r = .87$ and $.89$), agreement between independent sets of dollar juries is quite weak ($r = .42$). The variability of individual dollar judgments is so large that even the medians of 12 judgments are quite unstable. The problem could be reduced, of course, by taking larger samples. For example, we found that the median correlation between sets of dollar juries rises to $.80$ when the size of the juries is increased to 30 (median correlations for 30-person outrage and punishment juries rise to $.95$ and $.97$).

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	Predi Ratio
Joan	Large	High	\$500,000	\$2,000,000	\$15,000,000	5.14	
Joan	Medium	High	200,000	900,000	3,000,000	5.03	
Thomas	Medium	—	200,000	500,000	1,575,000	5.02	
Martin	Medium	High	350,000	1,000,000	4,000,000	4.98	
Thomas	Large	—	200,000	560,000	2,750,000	4.95	
Joan	Large	Low	175,000	1,000,000	12,500,000	4.93	
Martin	Large	High	350,000	1,900,000	10,000,000	4.92	
Frank	Medium	—	230,000	760,000	2,100,000	4.86	
Frank	Large	—	225,000	1,000,000	4,000,000	4.82	
Mary	Large	—	290,000	1,000,000	4,000,000	4.79	
Joan	Medium	Low	150,000	750,000	5,500,000	4.71	
Mary	Medium	—	250,000	710,000	2,100,000	4.70	
Martin	Large	Low	350,000	1,000,000	5,000,000	4.47	
Martin	Medium	Low	200,000	675,000	2,250,000	4.16	
Susan	Large	—	100,000	300,000	1,000,000	3.27	

Scenario	Firm Size	Harm Level	Lower 95% Confidence Bound	Median	Upper 95% Confidence Bound	Mean Jury Punishment	Predi Ratio
Susan	Medium	—	50,000	225,000	800,000	3.03	
Janet	Medium	High	100,000	200,000	690,000	2.79	
Carl	Medium	—	15,000	155,000	375,000	2.78	
Carl	Large	—	50,000	200,000	750,000	2.64	
Janet	Medium	Low	0	150,000	650,000	2.49	
Janet	Large	High	0	287,500	1,500,000	2.39	
Janet	Large	Low	12,500	200,000	1,000,000	2.38	
Jack	Large	High	0	0	350,000	1.24	
Jack	Medium	High	0	45,000	225,000	1.07	
Jack	Medium	Low	0	0	112,500	1.03	
Jack	Large	Low	0	2,550	500,000	0.95	
Sarah	Large	—	0	0	1,000	0.51	
Sarah	Medium	—	0	0	13,000	0.46	i

Median

Table 3. Median Correlations Between Sets of Simulated Juries

	<u>Outrage</u>	<u>Punishment</u>	<u>\$ Awards</u>
Outrage	.87		
Punishment	.86	.89	
\$ Awards	.47	.51	.42
Overall Median Award.	.71	.77	.69

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 28 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 preset conversion function, to generate punitive awards that would accurately represent community sentiment, thus reducing much of the unpredictability of awards.

E. The Underlying Problem: "Scaling without a Modulus"

A key to our analysis is a distinction that psychologists draw between two types of scales. (i) *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. (ii) *Magnitude scales* 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 response to socially relevant stimuli, such as the severity of crimes and the severity of punishments.¹⁰¹ 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 topic of much controversy,¹⁰² some characteristic differences between them are well established. (i) 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); (ii) judgments on magnitude scales are often erratic, in the sense of highly variable; (iii) 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

¹⁰¹ Stanley S. Stevens, *Psychophysics* (1975); see also note infra.

¹⁰² See, e.g., R. Duncan Luce and Carol Krumhansl, *Measurement, Scaling, and Psychophysics*, in *Stevens' Handbook of Experimental Psychology, Vol 1: Perception and Motivation* 3-74 (Richard Atkinson, Richard Herrnstein, Gardner Lindzey, and R Duncan Luce eds. 1988); L.E. Marks, *Magnitude Estimation and Sensory Matching*, 43 *Perception and Psychophysics* 511 (1988); Stanley S. Stevens, *Psychophysics* (1975).

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 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 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.¹⁰³ This reasoning is what led to the central hypothesis of the present study, a hypothesis that we described and established above.¹⁰⁴

¹⁰³ In the legal context, some moduli might even be insufficiently informative even if provided. If, for example, \$0 means entirely acceptable, and \$200 means objectionable (as in for example 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 managably organized in the likely comparison set.

¹⁰⁴ Similar evidence emerges from Michael Saks et al., *Reducing Variability in Civil Jury Awards*, 21 *Law and Human Behavior* 243 (1997). Saks et al. show that under experimental conditions, variability can be reduced in the context of pain and suffering awards by telling people average dollar awards for the type of injury at issue, intervals (where 80% of awards for similar injuries fell), average--plus-intervals, and examples (awards for four similar cases). *Id.* at 249. Without using the concepts, Saks et al. in effect supply a modulus in each of these conditions, and the result is dramatically to decrease variability *Id.* at 250-51. See also the

F. A simple way to improve predictability

Is it possible to improve the predictability of dollar awards? How might this be done? We performed a statistical analysis designed to answer these questions.

A conventional view about the role of juries in setting punitive awards is that the jury is a sample from the community, whose function it 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 12 citizens: use 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 test the effect of this procedure in our data, we first estimate the conversion function separately for medium and for large firms, since there was a significant effect of firm size on dollar awards. Following Stevens,¹⁰⁵ power functions were estimated, which related the mean punishment jury response for each case to the corresponding overall median of individual dollar awards, our measure of community sentiment. 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 transforming this value to dollars, separately for each jury, by the appropriate

proposal for a form of scaling, through baseline appraisals, in Glen O. Robinson and Kenneth S. Abraham, *Collective Justice in Tort Law*, 78 Va. L. Rev. 1481, 1490 (1992).

¹⁰⁵ See Chapman and Bornstein, *The More You Ask For, The More You Get*, 10 *Applied Cognitive Psych.* 519 (1996).

conversion function (depending on whether the defendant firm was large or medium-sized). To measure the accuracy of a simulated jury as an estimate of the population median, we 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. 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 28 cases in our study, one from dollar juries and one from punishment juries. In 25 of the 28 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 is interpreted to mean 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 estimated far more precisely, if the legal system used punishment juries and a conversion function rather than dollar juries. 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, plaintiff's demand, 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 below.

G. Context, Harm, Firm Size, and Other Findings

In this section we briefly report our findings on our other hypotheses, and discuss some issues of particular relevance to punitive damage reform.¹⁰⁶

1. The effect of context

Unlike real jurors, who are exposed to a single case for a long time, the participants in our study responded to a total of 10 product liability cases in quick succession, and had an opportunity to compare most of these scenarios to each other. 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. The experimental design provides a comparison of the distribution of judgments to each scenario when it is judged in isolation or in the context of other scenarios.

We examined whether the availability of a context of comparison affected the distribution of judgments; the question is important, since the legal system often forces juries to evaluate cases without a set of comparison. Our basic finding was that in a no-comparison condition, there is a cautious tendency to diminished differentiation in judgments about different cases. The availability of a context apparently makes a serious cases appear more serious than it would on its own, and makes a milder case appear milder. Thus the most consistent effect of a context of similar cases is to increase the *range* of the judgments across different scenarios. We conclude that the availability of a context of

¹⁰⁶ A detailed analysis can be found in Daniel Kahneman, David Schkade, and Cass R. Sunstein, *supra* note.

similar cases improves people's ability to discriminate among these cases, but does not affect the basic moral intuitions that the cases evoke. This point is connected to the topic of punitive damage reform below.

2. From outrage to punishment: the harm effect

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,¹⁰⁷ and we suspected that they would also be important to lay intuitions about the proper punishment for reprehensible actions. These predictions were tested by constructing alternative versions of four additional vignettes, which differed in the harm that the plaintiff had suffered. 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 less in the way of qualitative loss, as for example in the case of a burned child. (See the Appendix for examples.)

As predicted, we found that the degree of outrage evoked by the defendant's behavior was not affected by the harm that occurred. 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. As predicted by the outrage model, the significant harm effect found for punishment ratings carried through to dollar awards. Thus low harm produced an average award of \$727,599 and high harm an average award of a substantially greater amount: \$1,171,251.

¹⁰⁷ Thus there is a great deal of discussion 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).

3. *The effect of firm size*

Within the academic community, opinion is sharply divided on the question whether the amount of punitive awards should depend on the size of the defendant firm.¹⁰⁸ Lay intuitions, in contrast, are quite clear. A psychological analysis suggests that people think in terms of retribution rather than deterrence, and 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.¹⁰⁹) Our hypothesis was 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 firms (\$526,398). This is substantial evidence that equivalent outrage and punitive intent will produce significant higher dollar awards against wealthy defendants.

¹⁰⁸ See Polinsky & Shavell, *supra* note; Galanter and Luban, *supra* note; Kenneth Abraham and John Jeffries, Punitive Damages and the Rule of Law: The Role of Defendant's Wealth, 18 J. Legal Stud. 415, 415-19 (1989); Dan Dobbs, Ending Punishment in "Punitive" Damages, 40 Ala. L. Rev. 831, 871-72 (1989).

¹⁰⁹ 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, *supra* note, suggest that this view is implausible, but that question cannot be resolved a priori; it is an empirical issue. Compare Robert MacCoun, Differential Treatment of Corporate Defendants by Juries, 30 Law & Society Review 121, 133-39 (1996) (finding effects from corporate identity and commercial activity, but not from wealth per se).

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

A. General Observations, Anchoring, and a Roadmap

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 those 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 adversary 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 possibly reduce the degree of variance in real-world awards, at least if lawyers can anticipate which people will have outlier moduli. Nor did our “juries” deliberate.¹¹⁰ 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 did refer to compensatory damages and also to firm size, but it did not contain two usual “anchors”:¹¹¹ plaintiff’s demand and the jury’s own prior determination of compensatory damages. Such anchors are likely to matter a great deal to actual awards. There is experimental evidence that the plaintiff’s demand has considerable importance,¹¹²

¹¹⁰ We are now embarking on a follow-up study that does involve mock juries, and hence that attempts to be more precise about the effects of deliberation.

¹¹¹ With the exception of firm size and also compensatory damages, which might serve as an anchor, but probably less than in real world cases, whether those damages are chosen by deliberating juries and thus may have special salience.

¹¹² See Hastie, Schkade, and Payne (forthcoming).

and experimental evidence too of the effect of anchors in pain and suffering cases, which are analogous.¹¹³ There is real-world as well as experimental evidence of an anchoring effect from the compensatory award.¹¹⁴ 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.¹¹⁵ We suspect that people are not confident of the dollar amounts they award,¹¹⁶ and this lack of confidence increases the likely use of anchors. It is natural to think that the process of magnitude scaling without a modulus will encourage people to seize on whatever anchors are available, whether or not 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 less in the way of unpredictable outcomes. But if this is so, this particular form of predictability comes with a cost of its own: introducing an additional layer of arbitrariness, if (as if 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

¹¹³ See Michael Saks et al., *Reducing Variability in Civil Jury Awards*, 21 *Law and Human Behavior* 243, 254 (1997).

¹¹⁴ See Theodore Eisenberg et al., *supra* note.

¹¹⁵ See Karen E. Jacowitz and Daniel Kahneman, *Measures of Anchoring in Estimation Tasks*, 21 *Personality and Social Psychology Bulletin* 1165 (1995).

¹¹⁶ There is some support for this in our study; we asked people to say what they thought was the most difficult part of the task, and their answers emphasized the assignment of dollar amounts.

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 point to their apparently arbitrary character, and many proposed remedies attempt to promote rule of law values through, for example, more careful and more specific judicial instructions. In the *BMW* case, 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.¹¹⁷ 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 to juries. The problem is instead 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*.¹¹⁸

For purposes of obtaining the virtues associated with the rule of law, emphasized by Justice Breyer and many others, the solution lies in counteracting the arbitrariness that comes from the unbounded dollar scale of dollars. Of course rule of law virtues are not the only virtues of a legal system,¹¹⁹ and

¹¹⁷ 116 S. Ct. at 1605-1607.

¹¹⁸ There is the further problem that juries do not always follow even detailed instructions. See Hastie, Schkade, and Payne (forthcoming).

¹¹⁹ Cf. Joseph Raz, *The Rule of Law and Its Virtue*, in Joseph Raz, *The Autonomy of Law* (1986).

constraints on erratic awards may not produce all of the necessary reform. We will shortly turn to these issues.

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 those intuitions can be overcome in the courtroom. Perhaps deterrence plays some role in actual awards, 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.¹²⁰ 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,¹²¹ it does give

¹²⁰ See Polinsky and Shavell, and in particular the model jury instructions offered as Appendix A. 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 damage multiplier that corresponds to your estimated probability of escaping liability."

¹²¹ See Reid Hastie, David Schkade & John Payne (forthcoming), reporting the outcomes of mock jury tests of punitive damages. After deliberating during which the "jurors" received copies of the judge's instructions, participants were asked specific questions on the important elements on which they make judgments of punitive damages. A lenient

reason to questions 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).

Second, jurors are not likely to be good at the task of promoting optimal deterrence even if this is what they are seeking 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 have to master a high degree of technical knowledge about a wide variety of subjects. Hindsight bias will almost inevitably confuse the assessment:¹²² A jury is likely to find a bad outcome to have been likely to occur if it in fact occurred. Various heuristics and confusion will in all probability 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 of continued use of the jury involves the desire to elicit, and to make relevant for law, the community's judgment about

grading of this written, recall-comprehension test yielded a median score of only 5% correct.

Note also Anderson and MacCoun, Goal Conflict in Juror's Assessments of Compensatory and Punitive Damages (unpublished manuscript 1997), showing leakage between compensatory damages and punitive damages.

¹²² On the general topic of hindsight bias, see Baruch Fischhoff, Hindsight = Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty, 1 J of Experimental Psychology 288 (1975). See also Reid Hastie, David Schkade, and John Payne, (forthcoming 1997), which shows a hindsight bias directly for punitive damages—a difference, in an experimental setting, of a probability judgment of 37% without hindsight, and a probability judgment of 71% after hindsight.

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.¹²³ The task is to find a method to obtain that judgment without introducing arbitrariness and noise.

5. Isolating objections to the current system

Punitive damages reform should attempt to ensure that juries are charged with performing tasks that they can perform well, and should relieve juries from having to perform tasks that they perform poorly, in a way that produces excessive unpredictability,¹²⁴ confusion, and arbitrariness. And it would be reasonable to react to our study by suggesting a simple reform: Juries should decide questions of civil liability, just as

¹²³ See Galanter and Luban, *supra* note.

¹²⁴ Of course there is a question how bad it is for jury judgments to be unpredictable, and what kinds of unpredictability are 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 bad is that it may make planning more difficult if expected value is costly or impossible to calculate, and it may create overdeterrence in risk-averse actors; another reason is that it makes for a form of unfairness, 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.

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.¹²⁵ Of course there is a possible problem with judicial judgments about 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 antipopulist, for as we have seen, the award of any particular jury may well fail to reflect the community's sentiment on the topic of appropriate dollar award.

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 juries' dollar amounts, 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 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 the community as a whole about the appropriate dollar award*.

The second objection points to *defective calibration*, that is, to a poor translation of punitive intent into dollars. The problem is that juries lack the information that would enable them to undertake a good or accurate translation, since ordinary people cannot know the effects of a particular dollar award on a particular class of defendants.

¹²⁵ See Symposium, The Federal Sentencing Guidelines: Ten Years Later, 91 Nw. L. Rev. 1231 (1997), for various views on the guidelines now in place.

The third objection is directed against punitive intent and points to *improper grounds for judgment*. Here the complaint is that the jury is focusing on irrelevant factors, giving undue weight to relevant factors, giving insufficient weight to relevant factors, or even 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), a noise-free punitive damage judgment is objectionable too.

More particularly, 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 do is to show, on the basis of our study, how legal goals mesh, and fail to mesh, with an understanding of the psychological underpinnings of punitive awards. An important point in this connection is that ordinary people are intuitive retributivists, and there is a serious question about the role of retribution with respect to corporations and firms.¹²⁶ The retributive idea is most naturally and simply introduced with respect to people who have imposed harms;¹²⁷ 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

¹²⁶ See Polinsky and Shavell, *supra* note.

¹²⁷ See Jean Hampton, *supra* note, at 111.

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.

B. Punitive Damages Reform, I: Community Judgments about Dollars, or Predictable Populism

1. A modest proposal

From what we have said thus far, the most modest reform proposal is straightforward, and it is modest indeed. The goal of the modest reform is to get 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 come up with (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).

We have seen that if particular juries are asked to produce a dollar award, as an indicator of community sentiment, there will be a great deal of variability, and that 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 like that described above 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, 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.

2. 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.¹²⁸ There is also a question whether to alter the calibration function when social norms have changed. But these difficulties may not be insuperable; at 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.¹²⁹ The calibration could be done (say) every five years, to take account of changing social norms.

As Part III above shows, 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.

¹²⁸ We have dealt here with products liability cases; but is this 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?

¹²⁹ Some cautionary notes, in the context of pain and suffering, emerge from 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 Law Review* 1109, 1125 (1995), which discuss possible problems with jury understanding of cases.

3. *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 Phase One 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 necessary to compensate the plaintiff. The purpose of this punishment is to deter future wrongdoing and to reflect your view about the need just 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 an appropriate dollar amount 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."

4. *Ratings and rankings*

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 10.¹³⁰ The advantage of this approach is that it would lead to

¹³⁰ Relevant discussion and a related proposal can be found in David Baldus et al., *Improving Judicial Oversight of Jury Damages*

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. At least if they are accompanied only by verbal descriptions of the sort we have given here, the relevant numbers (0 to 6, 0 to 10) are likely to be perceived as having a highly artificial quality.¹³¹ In practice such scales can work quite well; consider the existence of popular and relatively informative movie ratings that use such scales, or ask whether it might be possible to rate 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. Perhaps most important, our data show that there is more (implicit) agreement on rankings than on absolute numbers. Hence the use of numbers will produce less predictability (though more precision) than an attempt to produce rankings, though it is still far better than dollars.¹³²

Assessments: A Proposal for the Comparative Additur/Remittitur Review of Awards for Nonpecuniary Harms and Punitive Damages, 80 Iowa Law Review 1109 (1995). The authors propose judicial review of awards by reference to comparison cases, building on existing practice. Our proposals represent a more dramatic departure from existing practice.

¹³¹ There is also a risk that a small set of numbers will make it difficult for the jury to make 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—for example, O.J. Simpson-style murders—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 make relevant discussions.

¹³² 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

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 very egregious wrongdoing, requiring a very substantial award.¹³³ The jury would be asked to assess the case at hand in terms of the sample cases, not with any dollar amounts or even numbers. 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

10, how was last night's dinner?) It is much easier to make comparisons—to compare meals, conferences, lecture, briefs, and the like against one another. Comparisons avoid the need to generate uniformity among heterogeneous people with respect to the meaning of the key terms.

¹³³ See the detailed and instructive discussion of comparison cases for additur and remittitur judgments, in Baldus et al., *supra* note, at 1153-1160; see also the discussion of pain and suffering schedules in Robert Bovbjerg et al., *Valuing Life and Limb in Tort: Scheduling "Pain and Suffering,"* 83 *NW 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 award in a systematic way. If the comparison cases are not so bad, it should 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. Mark Kelman, Yuval Rottenstreich, and Amos Tversky, *Context-Dependence in Legal Decisionmaking*, 25 *J Legal Stud.* 287, 288 (1996). 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. Tradeoff contrast arises when the option under consideration is evaluated more favorably in the presence of similar but clearly inferior options than it appears in the absence of such options. *Id.* at 288-289. (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.

pretest). The jury's task would be to compare the case at hand successively to each of the ten scenarios, to determine whether it is worse or less bad. The relevant score would be the number of scenarios that are better, plus 1. This gives you a ranking of the case in the set of 11, without requiring the jury to rank all 11. 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.¹³⁴ The jury could say, for example, that of the six cases provided, the case at hand is the next-to-worst, or closer to 5 than to 4. 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 cases, and also that it deserves more severe punishment than the most serious case.

A key advantage of this approach is that the jury would be selecting a dollar amount only implicitly, and not explicitly.¹³⁵ 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 with rankings, the jury will disagree with the legal system's rankings, and this may confound the whole

¹³⁴ Thus, for example, Baldus et al., *supra*, at 1154-1155, 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 patient available options, (3) delivery of nonapproved care, (4) intent to cause harm in delivery of care. There is a similar hierarchy for willful disregard. *Id.* What we are suggesting is that ideas of this kind might be used to develop scenarios for jury use.

¹³⁵ If the no-comparison effect had been stronger, this procedure would have been clearly superior. The limited effect of the no-comparison condition merely removes one reason for adopting the procedure, but it is still far better than the status quo, and seems better than any alternative that seeks to maintain a role for the jury in registering its moral judgment.

enterprise. It could be argued that the ranking might serve to educate jurors about community values and judgments, but it cannot be denied that some juries might reject what they are told. On this count ratings have an advantages. Another disadvantage of this 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.¹³⁶ Our own study of context provides additional support for the general idea of providing a context of cases, and the benefits would extend to both rankings and ratings.

5. 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.¹³⁷ Moreover, judicial review of punitive damage judgments tends to be highly individualized, and hence to rely on context-independent intuitions (putting to one side the fact that past experience may produce a sense of context and comparison).¹³⁸ Thus adjustments are often made by reference to the ratio between compensatory and punitive

¹³⁶ Michael Saks et al., *Reducing Variability in Civil Jury Awards*, 21 *Law and Human Behavior* 243, 250-51 (1997). 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.

¹³⁷ D.C. Barrett, Annotation, 15 A.L. R. 3d 1144 (1967).

¹³⁸ See Baldus et al., *supra* note, at 1132-1133.

damages (a crude test, for reasons suggested above) and ad hoc judgments about what seems “shocking.”¹³⁹

The idea of examining comparison cases is not, however, entirely foreign to the legal culture. Damage schedules and scaling through examples have been used successfully in the settlement of mass tort cases.¹⁴⁰ Such ideas have often been discussed in the context of pain and suffering awards.¹⁴¹ In a prominent case involving such awards, Judge Kearse, writing for the Second Circuit, attempted a careful comparison of the case at hand with twelve other cases.¹⁴² Nor has comparison been unavailable in the context of punitive damages, at least in the process of judicial review.¹⁴³ In an influential case, the court of appeals for the Ninth Circuit said that a district court should compare other punitive damage cases to a “a figure derived from the facts of the case at hand.”¹⁴⁴ And comparison of cases is a pervasive aspect of appellate review of punitive awards, furnishing a constraint on arbitrariness and inequality.¹⁴⁵ Of course the Supreme Court has not

¹³⁹ See, e.g., *Malandris v. Merrill Lynch*, 703 F.2d 1152, 1177 (10th Cir. 1981); *Baldus et al.*, *supra* note, at 1133.

¹⁴⁰ See Francis McGovern, *The Alabama DDT Settlement Fund*, 53 *Law & Contemp. Probs.* 61 (1990); Francis McGovern, *Resolving Mass Tort Litigation*, 69 *B. U. L. Rev.* 659 (1989).

¹⁴¹ David Baldus et al., *Improving Judicial Oversight of Jury Damages Assessments: A Proposal for the Comparative Additur/Remittur Review of Awards For Nonpecuniary Harms and Punitive Damages*, 80 *Iowa L. Rev.* 1109, 1122-1225, 1134-1137 (1995); Robert Bovbjerg et al., *Valuing Life and Limb in Tort: Scheduling “Pain and Suffering,”* 83 *NW L Rev* 908 (1989).

¹⁴² 748 F.2d 740 (2d Cir. 1980).

¹⁴³ See, e.g., *Bogan v. Stroud*, 958 F.2d 180, 186 (7th Cir. 1992); *Cash v. Beltmann N. Am. Co.*, 800 F.2d 109, 111 n. 3 (7th Cir. 1990); *Estate of Korf v. A. O. Smith*, 917 F.2d 480, 485 (10th Cir. 1990); *Ismail v. Cohen*, 899 F.2d 183, 186 (2d Cir. 1990); *Schultz v. Thomas*, 649 F. Supp. 620, 624-25 (E.D. Wis. 1986); *Sherrod v. Piedmont Aviation*, 516 F. Sup. 46, 56 (E. D. Tenn. 1978).

¹⁴⁴ *Morgan v. Woessner*, 997 F.2 1244, 1257 (9th Cir. 1993).

¹⁴⁵ See *Klein v. Grynberg*, 44 F. 3d 1497 (10th Cir. 1995); *Stafford v. Puro*, 63 F.3d 1436 (7th Cir. 1995); *Allahar v. Zahora*, 59 F.3d 693 (7th

insisted on this requirement as a matter of constitutional law. In an echo of some early death penalty cases,¹⁴⁶ the Court said that each case might be taken as *sui generis*. Because 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.”¹⁴⁷ Thus 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.¹⁴⁸

The Court’s reluctance to impose a constitutional requirement of comparing cases is understandable in light of the Court’s caution about proceeding at the constitutional level in the face of principles of federalism and gaps in the

Cir. 1995): *Ross v. Black and Decker, Inc.*, 977 F.2d 1178 (7th Cir. 1992); *Vasbinder v. Scott*, 976 F.2d 118 (2nd Cir. 1992); *Kimzey v. Wal-Mart Stores, Inc.*, 107 F.3d 568 (8th Cir. 1997); *Lee v. Edwards*, 101 F.3d 805 (2nd Cir. 1996); *King v. Macri*, 993 F.2d 294 (2nd Cir. 1993); *Michelson v. Hamada*, 36 Cal.Rptr.2d 343 (Cal. Ct. App. 1994); *Wollersheim v. Church of Scientology*, 6 Cal.Rptr.2d 532 (Cal. Ct. App. 1992); *Baume v. 212 E. 10 N.Y. Bar Ltd.*, 634 N.Y.S. 2d 478 (N.Y. App. Div. 1995); *Parkin v. Cornell University Inc.*, 581 N.Y.S.2d 914 (N.Y. App. Div. 1992):

¹⁴⁶ See *Woodson v. North Carolina*, 428 U.S. 280 (1976) (invalidating a mandatory death penalty statute).

¹⁴⁷ *TXO Production Corp. v. Alliance Resources*, 509 US 443, 457 (1993).

¹⁴⁸ Concurring, Justice Kennedy pointed to the high likelihood of legitimate inconsistency in jury results. Partly this is, in his view, a function of the fact that a jury is empaneled in a single case, not as a permanent body; partly this is a function of the generality of jury instructions. 509 U.S. at 472. Thus a lower court suggested that review of other jury determination “would undercut the jury system” and that various numbers are based on the fact that “juries hear unique facts and are given dissimilar instructions.” In *Re Exxon Valdez*, 1995 U.S. Dist. LEXIS 12952, 12959 n. 7 (D. Ala. 1995).

Court's knowledge of the actual world of punitive damage awards. Nothing said here demonstrates that such a requirement would be sensible 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 clear that a serious problem arises from the response mode of dollars and the difficulty of generating predictable dollar amounts. Our findings fortify the wisdom of the appellate practice of comparing punitive damage awards, and they suggest the possibility of a constitutional problem with awards that appear, in practice, to be stabs in the dark.

C. Punitive Damages Reform, II: Using Punitive Intent but not Dollars, or Technocratic Populism

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 abilities 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 involved, they are also using unbounded dollar scales, lacking a modulus that would give meaning to their estimates of different magnitudes. Hence it would be necessary to devise a translation formula that depends on the community's normative judgments about intended punishment, which are not only predictable but also worth using, but not on community judgments about dollar amounts, which can be made more predictable but which may not be worth using. To do this well, it would be desirable to translate the jury's intention to punish with expert assistance. Here are the two phases of a possible reform.

1. Phase one in the two-phase system: 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. The goal here too would be to obtain the jury's judgment about ratings or rankings. Thus 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.

2. Calibration and phase two: translating jury judgments into dollars

Phase two, we have noted, involves the translation of a jury's normative judgment—of intent to punish—into dollars. 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 dollar awards will actually do. In that

case, the point of a good translation is to ensure that the expressed punitive intention is turned into awards that are both predictable and faithful to what the jury truly intends to happen. 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, and are the people who suffer high-level officials, clerical workers, or consumers? The judgment about dollar awards would emerge through engagement with such issues.

D. Improper Grounds for Judgment and the Partial or Complete Elimination of Juries: Bureaucratic Rationality?

1. An insufficiently explored problem

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 this is the goal of punitive damages, surely it would be better for the judgment to be made by a judge or (better still) by a specialized regulatory agency entrusted with precisely that task.

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, 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 may want to punish a murderer via punitive damages even if the probability of detection was 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: high sensitivity to outrage (which probably implies low sensitivity to detection probability), substantial sensitivity to harm (which may not be fully expressed in the amount of compensatory damages), great sensitivity to firm size, and a retributive focus.

A decision to retain the jury system means 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 from a translation formula: juries may feel exposed to the possibility of being ignored, and may respond by behaving strategically. Thus there is a limit to the extent to which the translation formula can stray from common intuitions.

2. Improper grounds for judgment? Punitive damages reform, III

The most serious objection to our second proposal therefore comes from the view that juries do not and cannot easily be made to base their decisions on the proper grounds.¹⁴⁹ On this view, the problem is that the intention to punish, even if well-translated, is not an adequate way to assess punitive damage awards, because it is, from the

¹⁴⁹ See Hastie et al., *supra* note. 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. 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. *Id.*

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, it might be asked, wouldn't the same approach make sense for punitive awards?

The basic claim here would be that juries cannot be used to promote the aims, properly understood, of the system of punitive damages. There are 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 retributive motivations have played a large role in areas not involving juries. There is a continuing issue whether retributive judgments by juries can be sufficiently informed by relevant facts, and made reliably to express relevant values, so to serve as a sensible adjunct to the remaining fabric of the law. Even if we focus on retributive goals, can punitive damages be paid to fit with the rest of the law, so as to create a sensible regime of retribution?

The second objection would be 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 deterrent goals. 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 be better. Certainly it would seem that the inquiry into the probability of detection is 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 would be 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. Thus 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, 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 way our study’s basic point has been to give more precise content to this commonplace. In a nonpopulist republic that sometimes distrust 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 damage 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. An

administratively-operated schedule of fines and penalties would seem 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 designed substantially to reduce the jury’s role in the interest of more consistent and more expert judgments.¹⁵⁰ In the context of damages to natural resources, it has been suggested that contingent valuation should be replaced by a schedule of damages based on categories of harm;¹⁵¹ 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.¹⁵² In its current form, workers’ compensation attempts to deal with problems of valuation by placing a fixed dollar value on various injuries through a pre-determined schedule produced by a legislature or administrative agency.¹⁵³ The problem with the workers’ compensation system is its crudeness; it ignores

¹⁵⁰ See Baldus, *supra* note, at 1125-1131; Bovjberg et al., *supra* note, at 923

¹⁵¹ See Richard B. Stewart, *Damages to Natural Resources*, in *Analyzing Superfund 219*, 241-44 (Richard L. Revesz and Richard B. Stewart eds. 1995).

¹⁵² See Richard Epstein, *Cases and Materials on Torts* 1014-1038 (1994).

¹⁵³ *Id.* at 1035-1038.

possibly relevant individual variations. Its virtue lies in its speed, inexpensiveness, predictability, and consistency.¹⁵⁴

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.¹⁵⁵ The fact that judges, rather than juries, have traditionally made decisions about appropriate sentences raises a question about why a similar course is not followed for punitive damages.¹⁵⁶ The basic point is that in both of these cases, 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.¹⁵⁷ 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.

¹⁵⁴ See *id.*

¹⁵⁵ See Mashaw, *supra* note, at 150-165.

¹⁵⁶ 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 kinds of passion and zeal that might operate in a jury. That stigma may also make a degree of specialized experience, and even guidelines, especially important. And juries do, of course, have some control over sentencing through choices about criminal liability, especially when there are different theories of liability.

¹⁵⁷ See generally Jerry Mashaw, *Due Process in the Administrative State* (1984); see also Rubin, *supra*, discussing administrative fines (without, however, showing that the problems posed by scaling without a modulus have been overcome).

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, an issue to which we return below.

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, in order to ensure that public outrage plays a significant role in the legal system. It is also possible to fear that technocratic substitutes for the jury would 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.

E. Mixed Approaches, 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; its job might be to assess damages by comparing the case at hand to the preselected cases. This

approach would not take the whole subject of dollar awards away from the jury; it would attempt to root punitive awards in a set of antecedents judgments that could reasonably be compared with the case at hand. 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.¹⁵⁸ Doubtless a degree of 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, or require punitive damages to be within some multiple of compensatory damages, or allow judges to have a larger role in disciplining jury awards.¹⁵⁹ We might compare various reform proposals by asking which would contain 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 made to think in terms of optimal deterrence, both caps and multipliers are extremely crude. There is no reason to think 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

¹⁵⁸ See Michael Saks et al., *supra* note, at 149.

¹⁵⁹ See note *supra*.

deterrence. If the problem is that isolated juries come up with arbitrary or unpredictable outcomes, and if 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 respect to punitive intent, the conventional proposals are also far 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.¹⁶⁰ The best that can be said about a cap is that with little administrative cost, it will eliminate the most egregiously large judgments, but this virtue comes with many vices.

A damage multiplier might be a bit better, in the sense that it would also have low decision costs while allowing more flexibility than a cap by permitting very high awards when the compensatory damages are especially serious—while also having the advantage of preventing a jury from imposing unreasonable punishments on a individual or a corporation. But any multiplier would have crudeness of its own. No theory of punitive damages justifies a multiplier approach.

Of the proposed reforms, a shift from jury to judicial determinations of punitive damages appears to the most promising. Such a shift may well produce improvements over the current system, at least if it were thought that the relative populism of the jury allows illegitimate or irrelevant factors to

¹⁶⁰ See Michael Saks et al., *Reducing Variability in Civil Jury Awards*, 21 *Law and Human Behavior* 243 (1997) (finding that caps are likely to increase variability in the context of noneconomic compensatory losses). See also V.B. Hinsz and K. E. Indahl, *Assimilation To Anchors for Damage Awards in a Mock Civil Trial*, 25 *J. Applied Social Psych.* 991 (1995) (finding anchoring effect from caps).

play a role, or that a judge, because of her experience with multiple cases, can make more informed judgments. (Again compare criminal sentencing, of course undertaken by judges rather than juries.) Judges do reduce punitive awards that appear excessive, and because of their experiences judges have some comparative advantages;¹⁶¹ it is likely that a more general shift toward judicial control with reference to comparison cases would produce improvements. But judges are not likely to be able to capture the community's sentiments with respect to *either* dollars or punitive intent, and if the community's sentiments are irrelevant or only part of the appropriate inquiry, probably an administrative agency, with more detailed understanding of regulatory goals and instruments, should 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 hence while judicial judgments may reduce variance, there is likely to be a continuing problem of erratic judgments or use of anchors that introduce arbitrariness of their own. Judicially assessed punitive awards might well replicate some of the problems with judicially determined sentences. Thus 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.¹⁶² As we have suggested, various reform combinations and alternatives might be imagined, including dollar awards that are chosen after exposure to comparison cases.¹⁶³

We do not have sufficient information to evaluate all the possible alternatives here. But we can offer two general

¹⁶¹ See note *supra*.

¹⁶² See Baldus et al., *supra* note, at 1125-1131.

¹⁶³ Cf. Saks et al., *supra*, at 251 (discussing dollar awards in pain and suffering cases that are chosen after exposure to scales and similar disciplining devices).

conclusions. *First*, there is much to be said on the behalf of an incremental step, building on current practice: 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.¹⁶⁴ This incremental step would produce some of the gains sought by the first reform proposal discussed above. *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 based on a clear understanding of different forms of wrongdoing and of the consequences, for defendants, of different awards. Of course these specialists would 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.

Table 4 summarizes what has been a complicated discussion; our overall evaluations must be tentative because much depends on some unanswered empirical questions about the likely operation of the different systems.

F. A Note on the Constitution

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 v. Gore*? The most important point is that 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

¹⁶⁴ Compare Judge Weinstein's fascinating discussion of use of comparison cases in the context of pain and suffering, in *Geresy v. Digital Equipment*, 1997 U.S. Dist. LEXIS 14332 (Sept. 16, 1997).

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; but they do help to fortify the view that some awards should be taken to unacceptably arbitrary. We have shown that it cannot confidently be said that high jury awards, requiring payment of some amount $\$X$, reflect a well-considered community judgment in favor of $\$X$, or that jury encounters with the particulars of cases will lead to a situation in which particular facts are matched to dollar awards of X , $1/2$ of X , and $1/4$ of X . Indeed, our findings and our first reform proposal 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. As noted, moreover, we have found reason to doubt the suggestion that this problem can 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. Mostly the judicial role should come at the subconstitutional level, through review of punitive 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. None of our three proposals should raise constitutional problems. On the contrary, each of them attempts to overcome the difficulties to which the Court attempted to respond in *BMW v. Gore*.

Table 4. Punitive Damage Reform Possibilities

	Description	Analogies	Virtues	Vices	E
1. Current system	ad hoc judgments about particular events, with a populist aspiration	"Kadi 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 stand best if said is current are not
2. Caps	prevents the most excessive awards	current civil rights statutes, which also impose a cap	easy to administer, would prevent egregiously large awards	may increase variability; crudely tailored to any view of the problem or the purpose of punitive awards	unlike things and many ways 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 caps, better
4. Provide juries with other cases and their	retains jury authority over dollars	Proposals in the area of pain and suffering	Should improve predictability and increase rationality	Unclear how jury will respond to prior awards if it disagrees with	promise provide and judgment

	Description	Analogies	Virtues	Vices	E
accompanying punitive awards, in dollars		awards		them; unclear how to make sure the prior awards contain the right amounts	be toc
5. Strengthened or exclusive judicial control	“civil sentencing” model	current system of criminal justice	may reduce unpredictability, also produce more overall rationality	may also involve scaling without a modulus; eliminates or reduces populist elements; need to ensure judicial comparison of cases	mode: promi: to pro unpre irratio
6. Population-wide calibration function	a form of “predictable populism”	None	reduces variability, increases predictability	complicated administrative task	intrigu proba compl
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 either punitive intent of juries or the experts	intrigu promi: feasib above proba promi: below

	Description	Analogies	Virtues	Vices	E
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, most likely, as reduced costs

V. IMPLICATIONS, EXTENSIONS, SPECULATIONS

Our central finding—about the difficulty of mapping normative judgments onto an unbounded dollar scale—is relevant to a number of issues now facing law and policy. In this section, we briefly describe areas on which our data directly bear.

There is a unifying theme, having to do with the largely unexplored connection between actual or potential descriptive findings in psychology and the normative goals of the legal system. Suppose, for example, that an understanding of human 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. If this is true, the normative goals of law will be systematically frustrated. (To paraphrase Herbert Simon, people will not do what they cannot do.) At least to some extent, this appears to be so in the context of punitive damages; the psychology of “mapping” confounds some of the goals of the award of punitive damages. In the context of jury determinations, it would then be desirable to disentangle—as we have attempted to do here—the possible problems with the existing system, problems revealed as such by bringing normative analysis to bear on the descriptive account. Thus it may emerge that a single jury will deviate erratically from population-wide judgments, or that those judgments are flawed because they require judgments on intent to be mapped onto dollars, or that those judgments are flawed because they do not and cannot be made to grow out of what is, and should be, the basic goal of the legal system.

In many areas, an analysis of this basic sort may well apply. We do not know, for example, about the actual ingredients of a jury’s judgment that the appropriate compensatory damages for a libelous statement are, say, \$1 million—though we do have reason to think that outrage is highly relevant and that the compensatory award will be

inflated if punitive damages are unavailable.¹⁶⁵ In theory, of course, the law wants compensatory judgments in the area of libel to reflect the monetary value of what has been lost as a result of reputational harm. In practice, it is notoriously difficult to calculate that value—what evidence could reliably establish it?—and thus the law relies on crude surrogates. This leaves the following question: As a psychological matter, do juries do what they are supposed to do? Might they instead be making normative judgments about the outrageousness of the defendant's conduct, or the innocence of the plaintiff, and base libel awards on some combination of these judgments and arbitrary anchors? It is predictable, at least, that the psychology of libel awards and the goals of the legal system are in some tension. There is a large research agenda, and what we have done here is only a start. We outline a few possible areas for future investigation. We also 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. A basic underlying question has to do with 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.” Thus there are serious issues about the populist and technocratic dimensions of compensatory awards in these domains of the law.

¹⁶⁵ Anderson and MacCoun, Goal Conflict in Jurors' Assessments of Compensatory and Punitive Damages (unpublished manuscript 1997),

1. Pain and suffering

Awards for pain and suffering raise many of the same questions as punitive damages. To be sure, and importantly, 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, at the very least, hard to monetize. Thus a standard jury instruction says: “[T]he law allows you to award a plaintiff a sum that will reasonably compensate him for any past physical pain. . . . There are no objective guidelines by which you can measure the money equivalent of this type 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.”¹⁶⁶ An instruction of this kind offers little more guidance than a typical punitive damage instruction. What is the psychological process by which such awards are constructed? Can juries or judges make predictable or otherwise sensible judgments about dollar amounts? What are the ingredients of those judgments, whatever the instructions say? Perhaps most important: What does “compensate” mean, exactly?

Judgments about 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,¹⁶⁷ the resulting verdicts are notoriously variable, in a way that raises questions very much like those in the punitive damage setting.¹⁶⁸ In particular,

¹⁶⁶ G. Douthwaite, *Jury Instructions on Damages in Tort Actions* section 6-17, at 274 (2d ed. 1988).

¹⁶⁷ An effort at disciplining decision is made in W. Kip Viscusi, *Reforming Products Liability* 99-116 (1991)

¹⁶⁸ See G.A.O., *Medical Malpractice: Characteristics of Claims Closed* in 1984, 1-3, 18-19, 23-24, 40-42 (1987); Mark Geistfeld, *Placing A Price*

people with similar injuries are often awarded very different amounts of damages.¹⁶⁹ Studies have found that plaintiffs with relatively small losses tend to be overcompensated and those with large losses tend to be undercompensated, and also that there is a significant degree of randomness here.¹⁷⁰ 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 arbitrariness of their own. Thus 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 other aspects of the case as anchors.¹⁷¹ One study suggests that severity of injury explains only 40% of the variation in awards.¹⁷²

On Pain and Suffering, 83 Cal L Rev 773 (1995); Randall Bovbjerg et al., Valuing Life and Limb in Tort: Scheduling "Pain and Suffering," 83 NW L Rev 908 (1989); W. Kip Viscusi, Pain and Suffering in Product Liability Cases: Systematic Compensation of Capricious Awards, 8 Int'l Rev. of L. & Econ. 203, 204-208, 214-19 (1988).

¹⁶⁹ Saks et al., *supra* note, at 243; Geistfeld, *supra* note, at 784.

¹⁷⁰ Saks et al., *supra* note, at 245.

¹⁷¹ *Id.*

¹⁷² Bovbjerg et al. at 923. A recent study shows an additional point: Judgments about pain and suffering are highly sensitive to framing effects. Edward McCaffery, Matthew Spitzer, and Daniel Kahneman, Framing the Jury, 81 Va. L Rev 1341 (1995). 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 be paid in order to get 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 (a) 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 to (b) 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 recent study shows

There is the additional problem that 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 further 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, it will make sense to consider reforms of the sort discussed here. This could be done by moving in the direction of a damage schedule to cabin the jury's judgment¹⁷³ or by using a set of comparison cases for jury or judicial guidance.¹⁷⁴ As we shall shortly see, a 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.

substantial differences between (a) and (b), 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 "making whole" award. See *id* at 1388.

¹⁷³ See Geistfeld, *supra* note.

¹⁷⁴ See Patricia Danzon, *Medical Malpractice Liability*, in *Liability: Perspectives and Policy* 101, 121-24 (Robert E. Liton and Clifford Winston eds. 1988); Kenneth S. Abraham and Glen O. Robinson, *Aggregative Valuation of Mass Tort Claims*, 53 *Law & Contemp. Probs.* 137, 141-46 (1990); Peter Schuck, *Scheduled Damages and Insurance Contracts for Future Services*, 8 *Yale J. on Reg.* 213, 215-19 (1990); David Baldus et al., *Improving Judicial Oversight of Jury Damages Assessments: A Proposal for the Comparative Additur/Remittur Review of Awards For Nonpecuniary Harms and Punitive Damages*, 80 *Iowa L. Rev.* 1109, 1123(1995).

2. *Libel*

Similar issues arise in the law of libel, which notoriously lacks clear measures of damages.¹⁷⁵ In fact the common law rules governing libel reflect the difficulty of generating monetary amounts.¹⁷⁶ 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 easily correlated with monetary measures (“general damages”).¹⁷⁷ Sometimes plaintiffs are allowed to recover “presumed” general damages, that is, damages that are awarded without proving that they have suffered any actual damages, special or general.¹⁷⁸ 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? In practice, the resulting verdicts are unlikely to draw a sharp line between compensatory and punitive damages.¹⁷⁹

For present purposes let us notice that libel awards are likely to reflect effects similar to those we have discussed. Here the jury is likely 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.¹⁸⁰ A great

¹⁷⁵ See Richard Epstein, Was New York Times v. Sullivan Wrong?, 53 U Chi L Rev 782 (1986).

¹⁷⁶ *Id.*

¹⁷⁷ See Marc Franklin and David Anderson, Mass Media Law 214 (1995).

¹⁷⁸ *Id.*

¹⁷⁹ See Anderson and MacCoun, *supra* note x, showing “leakage” between punitive and compensatory awards: Where punitive awards cannot be provided, compensatory awards are higher.

¹⁸⁰ 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, Spitzer, and Kahneman, *supra* note.

deal of work remains to be done here as well, to understand the psychology of libel awards and their relationship to what the legal system is actually attempting to do.

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.¹⁸¹ 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? Damage caps, often proposed in the context,¹⁸² 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 distress¹⁸³ and sexual harassment.¹⁸⁴ Both of these are accompanied by damage remedies. With respect to such remedies, the basic story should be familiar. Monetization is

¹⁸¹ Franklin and Anderson, *supra* note, at 337, 342-46.

¹⁸² See William Van Alstyne, Reputation, Compensation, and Proof, 25 Wm. & Mary L. Rev. 793 (1984); Stanley Ingber, Rethinking Intangible Injuries, 73 Cal. L. Rev. 772 (1985).

¹⁸³ See Robert Keeton and William Prosser, The Law of Torts 54-66 (1984), for an overview.

¹⁸⁴ See Catharine MacKinnon, Sexual Harassment of Working Women (1979).

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

In the areas of intentional infliction of emotional distress and sexual harassment, there may well be a relatively uniform set of underlying judgments among different demographic groups, though with sexual harassment it would be most 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 vs. 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.¹⁸⁵ Punitive damages, by contrast, are intended to reflect a normative judgment about the

¹⁸⁵ Note in this regard the difference between the amount a plaintiff would require to deem himself restored, and the amount a plaintiff would demand to incur the injury in the first instance. See McCaffery, Spitzer, and Kahneman, *supra* note.

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 theory¹⁸⁶). At first glance this is a sharp distinction between the two. In this light would it make sense to consider reforms designed respectively to (a) capture a population-wide judgment about appropriate compensation, (b) capture a "compensatory intent" that would be mapped, by experts, onto dollar amounts, and (c) 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?

To answer this question it is necessary to ask why juries are now charged with the task of making judgments about appropriate compensation in cases in which that inquiry strains their factual 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."¹⁸⁷ Moreover, it is appropriate (on this view) to let the underlying decision reflect not merely facts but also the judgments of value that are held by the

¹⁸⁶ 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? 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 question, the jury's assessment inescapably creates normative issues. There is much room here for further descriptive work (what are the components of that assessment, in fact?) and normative work as well (what should the question of compensation be taken to mean in these various contexts?).

¹⁸⁷ See note *supra*.

community as a whole.¹⁸⁸ Whatever fact-finding 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, at least in these contexts, 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.

¹⁸⁸ An underlying question, in all of these areas, involves 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 once the defendant has been found liable, the defendant must take the plaintiff as the plaintiff experienced the injury; 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 also unclear that they should.

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 suggests grave reasons to doubt that this is in fact the case. Thus there is a serious question of reform strategies. How would the proposals discussed above work here? The first point to notice is 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, but far less useful, when punishment is not involved. To get a sense of the jury's numerical rating. A ranking might be used in various ways. 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"—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, rooted in a judgment of 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," 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 there the jury is not of course given a role, displaced as it is by an administrative law judge.¹⁸⁹

If the social security disability cases are really taken as a good analogy, technocratic considerations should predominate, and the third kind of reform proposal might seem 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.¹⁹⁰ 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.¹⁹¹ Whether it is desirable depends on the value of incorporating populist elements in the way that the more modest reforms promise to do.

¹⁸⁹ 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 have a sense of their practice, and to know whether they have created some of the same kind of variability discussed here. See Rubin, *supra* note, for a discussion of the similarity between punitive damages and administrative penalties.

¹⁹⁰ Compare the analogous proposal for additur and remittitur in Baldus et al., *supra* note.

¹⁹¹ 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. One issue here is how to minimize both decision costs and error costs, and in the abstract it is hard to know how much constraint on particularistic judgment will accomplish that task. For a good discussion, see Louis Kaplow, *Rules vs. Standards: An Economic Analysis*, 43 *Duke LJ* 557 *passim* (1992).

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.¹⁹² In view of the fact that similar problems beset other areas of the 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. 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.¹⁹³ 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).¹⁹⁴

Frequently, then, government must decide how much to expend per unit of regulatory benefit, and it is entirely to be expected that the phenomena that we have discussed here will come into play. Arbitrariness results partly from the

¹⁹² See Saks et al., *supra* note, at 246; Bovbjerg et al., *supra* note, at 953,

¹⁹³ See, e.g., Stephen Breyer, *Breaking the Vicious Circle* 1-15 (1993).

¹⁹⁴ See Tammy 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).

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 introduces a budget constraint, with accompanying comparative judgments, into governmental decisionmaking.¹⁹⁵ And Justice Breyer’s influential proposal of a special regulatory working group,¹⁹⁶ entrusted with the task of allocating resources to large problems rather than small ones, is of a piece with his concerns about rule-free punitive damage awards. The proposal is designed to ensure risk comparisons and to allow judgments about dollars 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, designed to ensure an institutional reform that captures the goal, appropriately understood, of risk regulation: extending human life.

Justice Breyer’s proposal has been criticized as excessively technocratic.¹⁹⁷ Perhaps risk regulation has multiple goals, and the extension of human life does not adequately capture them.¹⁹⁸ 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.¹⁹⁹

¹⁹⁵ See W. Kip Viscusi, *Fatal Tradeoffs* 42-49 (1992).

¹⁹⁶ See Stephen Breyer, *Breaking the Vicious Circle* 59-72 (1993).

¹⁹⁷ See Lisa Heinzerling, *Political Science*, 62 U. Chi. L. Rev. 258 (1995).

¹⁹⁸ *Id.*; Richard Pildes and Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. Chi. L. Rev. 1, 86-88 (1995)

¹⁹⁹ This is a goal of Richard Pildes and Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. Chi. L. Rev. 1, 86-95 (1995).

C. Mapping onto a Scale of Years

Our emphasis throughout has been on an unbounded dollar scale, but the legal system makes use of another scale: years. Criminal punishment of course 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 extreme ends; but it presents a similar difficulty of scaling without an obvious modulus.²⁰⁰ Thus the use of the scale of years presents some of the same questions as magnitude scaling in the absence of a modulus. Before the enactment of the Sentencing Guidelines, there were serious problems of arbitrary and unpredictable sentences,²⁰¹ 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 years of sentence? 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. Thus those who challenge the sentencing guidelines might be taken to be complaining, among other things, about the absence of an appropriate modulus around which to organize diverse sentences.²⁰²

²⁰⁰ See Stevens (1975) did magnitude scaling on jail sentences, and cross-modality matching against the severity of crimes. See also S.S. Stephens, (1966, *Science*, I believe) on 'scaling the social consensus' that was dedicated to that.

²⁰¹ See Jon Newman, *Federal Sentencing Guidelines: A Risk Worth Taking*, Brookings Rev. 29 (Summer 1987); Sandra Shane-DuBow, Alice P. Brown, and Erik Olsen, *Sentencing Reform in the United States* (U.S. Dept. of Justice, National Institute of Justice 1985).

²⁰² See generally Albert W. Alschuler, *The Failure of Sentencing*

D. A Note on 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.²⁰³ This problem is well worth considering, both because it is of interest in its own right and because it is closely related to the issue of punitive damages in the particular sense that it involves the mapping of a kind of attitude, desire, or judgment onto an unbounded dollar scale.²⁰⁴

The goal of contingent valuation methods is of course to decide how much to value goods that are not traded on markets. Some people think that valuation might be ascertained by looking at how people value goods that are in fact traded on markets.²⁰⁵ 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 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.²⁰⁶

Despite their apparent promise, contingent valuation methods have serious limitations, involving the difficulty of mapping normative judgments onto dollars and the problems

Guidelines: A Plea for Less Aggregation, 58 U. Chi. L. Rev. 901 (1991).

²⁰³ See the various perspectives in Symposium, 8 J. Econ. Persp. 3 (1994).

²⁰⁴ There is also a question whether it is sensible to assume that people have well-formed judgments on such questions.

²⁰⁵ See Viscusi, *supra*.

²⁰⁶ See William H. Desvousges et al., *Measuring Non-Use Damages Using Contingent Valuation*, Research Triangle Monograph 92-1; George Tolley et al., *Valuing Health for Policy* 290-94 (1994); Symposium, *Contingent Valuation*, J Econ Persp 3 (Fall 1994).

created by framing effects. A special problem is that of indifference to quantity, or inadequate sensitivity to scope, reflected in the fact that people will give the same dollar number to save 2000, 20,000, and 200,000 birds—or the same number to save one, two, or three wilderness areas.²⁰⁷ 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 as they are willing to pay to maintain fishing in all Ontario lakes.²⁰⁸ Thus a similar WTP was found to preserve 110 or 10,000 acres of wetland in New Jersey.²⁰⁹ Relatedly, 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 seems crucial in determining valuation. When asked for their willingness to pay to preserve visibility in the Grand Canyon, people offer a number five times higher when this is the first question than when it is the third question.²¹⁰

What unifies contingent valuation and punitive damage assessment is the problem of mapping a normative judgment onto an unbounded dollar scale. Let us notice, then, some similarities and differences between contingent valuation and punitive damages. CV studies are sometimes understood as a method of ascertaining private willingness to pay for public goods. Despite the flaws of the method if so seen, contingent valuation is intended as a substitute for market measures.²¹¹

²⁰⁷ See Peter Diamond & Richard Hausman, *Contingent Valuation*, 8 *J Econ Persp* 45 (1994); Daniel Kahneman & Iliana Ritov, *Determinants of Stated Willingness to Pay*, 9 *J Risk & Uncertainty* 5 (1994).

²⁰⁸ Daniel Kahneman and Jack Knetch, *Valuing Public Goods*, 22 *J Env Econ and Management* 57 (1992).

²⁰⁹ William H. Desvousges et al., *Measuring Non-Use Damages Using Contingent Valuation*, *Research Triangle Monograph* 92-1.

²¹⁰ See *id.*

²¹¹ Of course a large question is whether the answers received in the contingent valuation setting do reflect private willingness to pay in the

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 the acontextual character of isolated judgments—how much is a certain species worth? how bad was an oil spill?—should make individual judgments less reliable.²¹² In fact our findings here have a parallel in the key conclusions of a study of contingent valuation,²¹³ 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 WTP, and confirmed that the amount of systematic variance was much lower for WTP than for these scales. Furthermore, transformations to logs or ranks improve things there, just as they do here.²¹⁴

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

market sense or instead sometimes else.

²¹² Judgments are also sensitive to framing effects. When asked how much they are willing to pay to allow a species to be lost, they 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* 167-177 (1994).

²¹³ See Daniel Kahneman & Iliana Ritov, *Determinants of Stated Willingness to Pay*, 9 *J Risk & Uncertainty* 5 (1994).

²¹⁴ *Id.*

have discussed here.²¹⁵ 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 in the original scale.²¹⁶ Rather than asking people 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;²¹⁷ 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 may 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 in making judgments on an unbounded scale of dollars helps fortify and specify the basis for Justice Breyer’s complaint,²¹⁸ grounded in rule of law

²¹⁵ See Daniel Kahneman & Iana Ritov, Determinants of Stated Willingness to Pay, 9 *J Risk & Uncertainty* 5, 29-30 (1994).

²¹⁶ *Id.* at 30.

²¹⁷ For further discussions see citations in notes *infra*.

²¹⁸ See *BMW of North America v. Gore*, 116 S. Ct. 1589, 1604-1609 (1996) (Breyer, J., concurring).

considerations, that some jurisdictions do not provide sufficient constraints on jury discretion. 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, or work against the jury's determination. We have suggested 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. 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, of rejecting punitive awards that are out of line with general practice and relevant

comparison cases²¹⁹—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—probably 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 compensatory damage awards, or punitive damage awards, for pain and suffering or sexual harassment; it would be valuable too 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

²¹⁹ See *Klein v. Grynberg*, 44 F. 3d 1497 (10th Cir. 1995); *Stafford v. Puro*, 63 F.3d 1436 (7th Cir. 1995); *Allahar v. Zahora*, 59 F.3d 693 (7th Cir. 1995); *Ross v. Black and Decker, Inc.*, 977 F.2d 1178 (7th Cir. 1992); *Vasbinder v. Scott*, 976 F.2d 118 (2nd Cir. 1992); *Kimzey v. Wal-Mart Stores, Inc.*, 107 F.3d 568 (8th Cir. 1997); *Lee v. Edwards*, 101 F.3d 805 (2nd Cir. 1996); *King v. Macri*, 993 F.2d 294 (2nd Cir. 1993); *Michelson v. Hamada*, 36 Cal.Rptr.2d 343 (Cal. Ct. App. 1994); *Wollersheim v. Church of Scientology*, 6 Cal.Rptr.2d 532 (Cal. Ct. App. 1992); *Baume v. 212 E. 10 N.Y. Bar Ltd.*, 634 N.Y.S. 2d 478 (N.Y. App. Div. 1995); *Parkin v. Cornell University Inc.*, 581 N.Y.S.2d 914 (N.Y. App. Div. 1992). 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.

awards in cases in which market measures are unavailable, and it throws into sharp relief the question whether this task should not 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 evaluated by an assessment of the appropriate place of purely factual judgments, for which juries are not especially well-suited. Thus the three kinds of punitive damage reforms may well have parallels in each of these areas.

Third, 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.²²⁰ In their various social roles, people lack a preexisting “master list” of values and preferences from which to make selections in situations of choice. This is true for consumer choices, which depend on context and alternative options, as well as for medical decisions,²²¹ politics, and

²²⁰ See Amos Tversky, Rational Theory and Constructive Choice, in *The Rational Foundations of Economic Behavior* 185 (Kenneth Arrow et al. eds. 1997).

²²¹ See A. Bastardi and Eldar Shafir, On the search for and misuse of useless information (forthcoming 1997).

law.²²² It is true as well for judgments made by jurors, judges, representatives, and citizens generally. Rather than emerging from some menu in the mind, these judgments are a function of procedure, 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 they are ill-equipped—or, more specifically, 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.

²²² See Mark Kelman, Yuval Rottenstreich, and Amos Tversky, Context-Dependence in Legal Decisionmaking, 25 J Legal Stud. 287 (1996).

APPENDIX 1. 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 law suits 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 fully compensate 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 2. 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 work place 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 pre-market tests, the defect appeared on several occasions. But the pre-market tests were not extensive, despite the fact that auto industry regulations require elaborate testing. Internal company documents show National Motors’ 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 were 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 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 non-prescription 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 non-prescription 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 others 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, “the federal government has many other things to worry about and probably won't bother us on this” and in any case “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 non-prescription 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 are 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

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 3. RESPONSE MODE MANIPULATION

Outrage

Which of the following best expresses your opinion of the defendant's actions? (please circle your answer)

Completely Acceptable	Objectionable	Shocking	Absolutely 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)

\$ _____

APPENDIX 4: EXPERIMENT DESIGN

Samples and Procedures

The sample had good representation from various demographic and socio-economic groups. For example, respondents were 44% male; 64% Caucasian, 16% African-American, 15% Hispanic; median income = \$30K-\$50K; median education = some college; median age = 30-39. 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. Participants were run in large groups at pre-arranged times over a four day period. Most respondents completed their task in 30 to 45 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 question and debriefing information. The instructions (which are excerpted in Appendix 1) 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).

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.

In Parts 1 and 2, between-subjects manipulations were response mode (Outrage, Punishment or \$ Damages) and firm size (annual profits of \$10-\$20 million (Medium) or \$100-\$200 million (Large)), and scenario sequence, including which scenario was evaluated first, in isolation from the others (in Part 1 of the procedure). Scenario order was counterbalanced using a Latin square so that each scenario appeared in each ordinal position with equal frequency. Table 2 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 “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).

This Working Paper is a tentative and preliminary version of an article that will be published in the Yale Law Journal. The research on which the article is based will be described in detail in a forthcoming article by Daniel Kahneman, David Schkade, and Cass R. Sunstein in the Journal of Risk and Uncertainty. Readers with comments should address them to:

Cass R. Sunstein
Karl N. Llewellyn Distinguished Service Professor of
Jurisprudence
The Law School
The University of Chicago

1111 E. 60th Street
Chicago, IL 60637

Chicago Working Papers in Law and Economics
(Second Series)

1. William M. Landes, Copyright Protection of Letters, Diaries and Other Unpublished Works: An Economic Approach (July 1991).
2. Richard A. Epstein, The Path to *The T. J. Hooper*: The Theory and History of Custom in the Law of Tort (August 1991).
3. Cass R. Sunstein, On Property and Constitutionalism (September 1991).
4. Richard A. Posner, Blackmail, Privacy, and Freedom of Contract (February 1992).
5. Randal C. Picker, Security Interests, Misbehavior, and Common Pools (February 1992).
6. Tomas J. Philipson & Richard A. Posner, Optimal Regulation of AIDS (April 1992).
7. Douglas G. Baird, Revisiting Auctions in Chapter 11 (April 1992).
8. William M. Landes, Sequential versus Unitary Trials: An Economic Analysis (July 1992).
9. William M. Landes & Richard A. Posner, The Influence of Economics on Law: A Quantitative Study (August 1992).
10. Alan O. Sykes, The Welfare Economics of Immigration Law: A Theoretical Survey With An Analysis of U.S. Policy (September 1992).
11. Douglas G. Baird, 1992 Katz Lecture: Reconstructing Contracts (November 1992).
12. Gary S. Becker, The Economic Way of Looking at Life (January 1993).
13. J. Mark Ramseyer, Credibly Committing to Efficiency Wages: Cotton Spinning Cartels in Imperial Japan (March 1993).
14. Cass R. Sunstein, Endogenous Preferences, Environmental Law (April 1993).

15. Richard A. Posner, What Do Judges and Justices Maximize? (The Same Thing Everyone Else Does) (April 1993).
16. Lucian Arye Bebchuk and Randal C. Picker, Bankruptcy Rules, Managerial Entrenchment, and Firm-Specific Human Capital (August 1993).
17. J. Mark Ramseyer, Explicit Reasons for Implicit Contracts: The Legal Logic to the Japanese Main Bank System (August 1993).
18. William M. Landes and Richard A. Posner, The Economics of Anticipatory Adjudication (September 1993).
19. Kenneth W. Dam, The Economic Underpinnings of Patent Law (September 1993).
20. Alan O. Sykes, An Introduction to Regression Analysis (October 1993).
21. Richard A. Epstein, The Ubiquity of the Benefit Principle (March 1994).
22. Randal C. Picker, An Introduction to Game Theory and the Law (June 1994).
23. William M. Landes, Counterclaims: An Economic Analysis (June 1994).
24. J. Mark Ramseyer, The Market for Children: Evidence from Early Modern Japan (August 1994).
25. Robert H. Gertner and Geoffrey P. Miller, Settlement Escrows (August 1994).
26. Kenneth W. Dam, Some Economic Considerations in the Intellectual Property Protection of Software (August 1994).
27. Cass R. Sunstein, Rules and Rulelessness, (October 1994).
28. David Friedman, More Justice for Less Money: A Step Beyond *Cimino* (December 1994).
29. Daniel Shaviro, Budget Deficits and the Intergenerational Distribution of Lifetime Consumption (January 1995).
30. Douglas G. Baird, The Law and Economics of Contract Damages (February 1995).

31. Daniel Kessler, Thomas Meites, and Geoffrey P. Miller, Explaining Deviations from the Fifty Percent Rule: A Multimodal Approach to the Selection of Cases for Litigation (March 1995).
32. Geoffrey P. Miller, Das Kapital: Solvency Regulation of the American Business Enterprise (April 1995).
33. Richard Craswell, Freedom of Contract (August 1995).
34. J. Mark Ramseyer, Public Choice (November 1995).
35. Kenneth W. Dam, Intellectual Property in an Age of Software and Biotechnology (November 1995).
36. Cass R. Sunstein, Social Norms and Social Roles (January 1996).
37. J. Mark Ramseyer and Eric B. Rasmusen, Judicial Independence in Civil Law Regimes: Econometrics from Japan (January 1996).
38. Richard A. Epstein, Transaction Costs and Property Rights: Or Do Good Fences Make Good Neighbors? (March 1996).
39. Cass R. Sunstein, The Cost-Benefit State (May 1996).
40. William M. Landes and Richard A. Posner, The Economics of Legal Disputes Over the Ownership of Works of Art and Other Collectibles (July 1996).
41. John R. Lott, Jr. and David B. Mustard, Crime, Deterrence, and Right-to-Carry Concealed Handguns (August 1996).
42. Cass R. Sunstein, Health-Health Tradeoffs (September 1996).
43. Douglas G. Baird, The Hidden Virtues of Chapter 11: An Overview of the Law and Economics of Financially Distressed Firms (March 1997).
44. Richard A. Posner, Community, Wealth, and Equality (March 1997).
45. William M. Landes, The Art of Law and Economics: An Autobiographical Essay (March 1997).
46. Cass R. Sunstein, Behavioral Analysis of Law (April 1997).

47. John R. Lott, Jr. and Kermit Daniel, Term Limits and Electoral Competitiveness: Evidence from California's State Legislative Races (May 1997).
48. Randal C. Picker, Simple Games in a Complex World: A Generative Approach to the Adoption of Norms (June 1997).
49. Richard A. Epstein, Contracts Small and Contracts Large: Contract law Through the Lens of Laissez-Faire (August 1997).
50. Cass R. Sunstein, Daniel Kahneman, and David Schkade, Assessing Punitive Damages (with Notes on Cognition and Valuation in Law) (December 1997)