

1999

Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages

Cass R. Sunstein

David Schkade

Daniel Kahneman

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



Part of the [Law Commons](#)

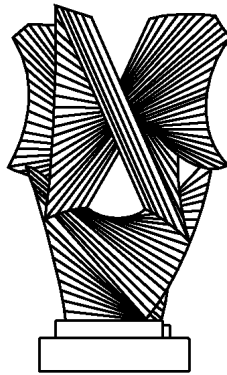
Recommended Citation

Cass R. Sunstein, David Schkade & Daniel Kahneman, "Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages" (John M. Olin Program in Law and Economics Working Paper No. 81, 1999).

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.

CHICAGO

JOHN M. OLIN LAW & ECONOMICS WORKING PAPER NO. 81
(2D SERIES)



Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages

David Schkade, Cass R. Sunstein, and Daniel Kahneman

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

This paper can be downloaded without charge at:

The Chicago Working Paper Series Index:
<http://www.law.uchicago.edu/Publications/Working/index.html>

The Social Science Research Network Electronic Paper Collection:
http://papers.ssrn.com/paper.taf?abstract_id=177368

Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages

David Schkade, Cass R. Sunstein,** and Daniel Kahneman****

I. Introduction

How does group deliberation affect individual judgments? How does the outcome of jury deliberations differ from some aggregation of individual decisions pre-deliberation? Speculation is not difficult. Perhaps juries converge toward the mean of individual judgments; perhaps juries move away from, or toward, the high or low of individual extremes. Perhaps juries approach an outcome that is more just or more accurate; perhaps juries produce more predictable and less erratic judgments, so that unpredictability at the individual level, or at the level of the mean or median of (six or twelve) individual judgments, does not exist at the jury level. A pervasive question is whether a deliberating jury has the effect of producing outcomes that treat the similarly situated similarly—perhaps in terms of civil or criminal liability (do people who have engaged in the same conduct receive the same verdict?), perhaps in the determination of appropriate damage awards, either compensatory or punitive (do similarly situated people receive the same awards?).¹

We attempt to make some progress on these questions. We do so principally by reporting the results of an extensive study of mock

* Professor of Management and William R. Spriegel Fellow, Graduate School of Business, University of Texas, Austin.

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

*** Eugene Higgins Professor of Psychology and Professor of Public Affairs, Princeton University. The authors are grateful to Exxon Company, U.S.A. for support of the research in this Article. Exxon bears no responsibility for our analysis or our conclusions; the data reported and the opinions expressed here belong to the authors. For helpful comments we are grateful to Robert MacCoun, Eric Posner, and Richard Posner.

¹ For concerns along this line, see *BMW of North America v. Gore*, 116 S Ct 1589 (1996); John Calvin Jeffries, Jr., *A Comment on the Constitutionality of Punitive Damages*, 72 Va L Rev 139 (1986).

juries (over 500 juries in total). In the study, these juries were asked to deliberate about punitive damage awards; individual judgments were collected before the jury deliberated. To compress a long story, our principal finding is that *deliberation increased both unpredictability and variance, by making large awards much larger and small awards smaller still*. Thus the principal effect of deliberation is often to polarize individual judgments, a pattern that has been found in many other group decision making contexts.² This finding—the first of its kind in the particular context of punitive damage awards—has important implications for jury awards involving both punitive and compensatory damages. It bears on group deliberation in other contexts as well. The same phenomenon of polarization that occurs in our punitive damage juries can be used to explain why a group of people moderately predisposed in favor of gun control will, after discussion, tend to be more than moderately predisposed in that direction, and why a group of individuals cautiously opposed to affirmative action is likely, after discussion, to oppose affirmative action with considerable fervor.

The study reported here has the advantage of being extremely close to—in fact part of the design is based on—an earlier one involving not deliberating juries but responses of 899 individuals to punitive damage cases.³ Our earlier study showed a remarkable consensus in individual judgments, made along a bounded numerical scale, about a series of personal injury cases. That study therefore found that *with respect to the underlying moral evaluation, different (synthetic, non-deliberating⁴) juries are likely to reach similar*

² See, e.g., David Myers and Helmut Lamm, The Group Polarization Phenomenon, 83 Psych. Bulletin 602 (1976); Daniel Isenberg, Group Polarization: A Critical Review and Meta-Analysis, 50 J. Personality and Social Psychology 1141 (1986).

³ See Cass R. Sunstein, Daniel Kahneman, and David Schkade, Assessing Punitive Damages, 107 Yale LJ 2071 (1998). This presentation is geared to analysis of the legal issues; the underlying data, and relevant psychological points, are presented in more detail in Daniel Kahneman, David Schkade, and Cass R. Sunstein, Shared Outrage and Erratic Awards: The Psychology of Punitive Damages, 16 Journal of Risk and Uncertainty 49 (1998).

⁴ As explained below, synthetic juries consisted of random groups of twelve individual judgments, with the mean or median judgment of the group of twelve reflecting the “verdict.” See 107 Yale LJ at 2098.

*conclusions about the relative severity of different cases.*⁵ Thus all-white, all-poor, all-rich, all-educated, all poorly educated, all-male, all-female, all-young, and all-old juries would probably come to very similar rankings of a set of cases, at least in personal injury cases and very possibly elsewhere.⁶ It follows that averaging the judgments of any random group of twelve people is likely to produce a moral judgment that predicts, with a reasonably high degree of accuracy, the judgment of any other group of twelve people. At the same time, the study found that *when asked to assess cases in terms of dollars, (synthetic, non-deliberating) juries become extremely unpredictable.*⁷ They become unpredictable in the specific sense that the judgment of any particular group of twelve is a poor predictor of the judgment of other groups of twelve.⁸ We speculated that it is the dollar scale that accounts for evidently erratic monetary judgments in many areas of the law, including not only punitive damages but compensatory awards in areas involving libel, sexual harassment, pain and suffering,⁹ and intentional infliction of emotional distress.¹⁰ When juries use the dollar scale, their verdicts probably do not treat the similarly situated similarly, with some awards that are arbitrarily high and others that are arbitrarily low. This unpredictability is also likely to produce overdeterrence in risk-averse defendants or in any case muffled and confusing (and to that extent costly) signals.¹¹

A natural question is, however, raised by these conclusions. This question is one of concern to those interested in damages, juries, and

⁵ Id. at 2095-2100.

⁶ Id. at 2097-2100.

⁷ Id. at 2100-2104.

⁸ Note that in our study we hold constant several factors that can be used to capture some of the variability in punitive damage awards, such as compensatory damages, case category, case particulars, and jurisdiction. It has been proposed by some authors that when analyzed using these factors, punitive awards are reasonably predictable. See Theodore Eisenberg et al., *The Predictability of Punitive Damages*, 26 J. Legal Stud. 623 (1997). Because we hold these factors constant, the unpredictability that we documented previously, and that we document here, cannot be accounted for by any of these factors.

⁹ See, e.g., David Leebron, *Final Moments: Damages for Pain and Suffering Prior to Death*, 64 N.Y.U. L. Rev. 256 (1989).

¹⁰ Sunstein, Kahneman, and Schkade, *supra*, at 2131-2140.

¹¹ See Paul Rubin et al., *BMW v. Gore: Mitigating the Punitive Economics of Punitive Damages*, 1997 Sup. Ct. Econ. Rev. 179, 184.

deliberative processes in general¹²: whether a deliberating jury might reduce or even eliminate the unpredictability of dollar awards. How does the process of jury deliberation affect the remarkable moral consensus? Will different juries converge toward the same dollar amount? The answers should have implications not only for punitive awards, but also for other damage judgments, certainly when these are hard to monetize, and possibly for questions of civil and criminal liability as well and even for deliberation generally. An additional question, also important, is what happens to dollar awards when juries deliberate: As compared with individual conclusions pre-deliberation, do such awards tend to go up, go down, or remain the same?

On the basis of a study of 3048 jury-eligible citizens in Phoenix, Arizona, participating in 509 mock juries, we offer two principal conclusions here, which we now state in somewhat less compressed form. The first is that *juries do not produce less erratic and more predictable awards than individuals*. On the contrary, deliberating juries tend to generate, with respect to dollar awards, even greater variance and unpredictability than is present in the judgments of the jurors who compose them. This is no mere technical conclusion; it suggests that the problem of erratic awards is intensified, not reduced, by deliberating juries.

The second conclusion, a corollary of the first, is *that deliberating juries significantly, and systematically, increase large punitive damage awards and also decrease small punitive damage awards*. In our study, jury judgments were more lenient than individual juror judgments for cases of less serious misconduct by the defendant; for cases of moderately serious and very serious misconduct, however, jury judgments were harsher, in many cases dramatically so. In 21% of the cases, the jury verdict was as high as or higher than that of the highest individual judgment pre-deliberation. As we discuss below, this result is part of the general phenomenon of *group polarization*, in accordance with which a deliberating group tends to shift toward a more extreme version of

¹² On the latter topic, see Amy Gutmann and Dennis Thompson, *Democracy and Disagreement* (1997); our analysis of group polarization raises some questions for the deliberative conception of democracy, though we leave those questions largely implicit here.

the views its members brought with them. The phenomenon of group polarization—with discussion amplifying initial individual tendencies, whatever they may be¹³—operates in accordance with identifiable mechanisms (thus far unexamined within the legal culture), which, unfortunately, do not suggest that the shifts are necessarily desirable, or that they implement any legitimate concerns of the legal system.

What follows at the normative level? The safest and most cautious conclusion is that to the extent that unpredictable punitive damage awards raise a serious concern, the problem is magnified rather than diminished by the process of deliberation. But is the increase in awards a tribute to the benefits of collective discussion? Without an independent theory of what awards should be, the evidence found here does not rule out that hopeful conclusion. But it certainly provides no affirmative support for it, or for the suggestion that group deliberation will, in this context, increase the rationality and soundness of outcomes. On the contrary, the substantial increases of high awards and the (smaller) decreases in low awards are predictable consequences of the effects of deliberation on a scale having a lower bound but lacking an upper bound.

In these circumstances, there are two implications for reform. The first involves the possibility, now shown to be feasible, of asking jurors to generate not a dollar figure, but a normative judgment on a bounded numerical scale. As we will explain, the deliberating juries in this study were able to use a punishment scale quite effectively; such a scale might be used as the foundation of dollar awards, in a way that could dramatically decrease unpredictability. The second implication follows from the fact that deliberation increases high awards and decreases low ones, in a way that increases variance and unpredictability. To the extent that unpredictability is a problem, our findings provide additional support for the view that punitive awards should be made not by juries, but by judges or some kind of administrative institution.¹⁴

¹³ Note also that mere exposure to the views of others, without discussion, has a similar amplifying effect. See David Myers, *Polarizing Effects of Social Comparison*, 14 *J Experimental Soc. Psych.* 554 (1978).

¹⁴ See, e.g., W. Kip Viscusi, *Stanford Law Review* (forthcoming 2000). The point is generally discussed in Sunstein, Kahneman, and Schkade, *supra* note, at 2126-2130.

This Article comes in six parts. Part II sets the stage, with brief discussions of the problem of calculating punitive damage awards, the legal background, the existing literature on unpredictable awards, and the effects of deliberation. Part III, the heart of the Article, discusses the methodology and results of the current study. Part IV investigates group polarization and associated causal mechanisms, involving social influences and persuasive arguments. Part V discusses the implications of the study for punitive damage reform. Part VI is the conclusion.

II. Theoretical Preliminaries: Outrage and Scales

Jury awards of punitive damages have become one of the most controversial topics in modern public law.¹⁵ To take just one example, an award of \$4.9 billion against General Motors attracted a great deal of national attention in July, 1999.¹⁶ It is now clear that the due process clause imposes constraints on permissible awards.¹⁷ A number of statutes, enacted and proposed, create punitive damage “caps,”¹⁸ and high awards have become a primary impetus for tort reform in general.¹⁹ There are also controversial issues about punitive damage awards in civil rights cases, most notably sexual harassment.²⁰ At the same time, the problems created by punitive awards bear on related questions in other areas of the law, involving,

¹⁵ See, e.g., Symposium, 1998 Wisconsin L. Rev. 1; A. Mitchell Polinsky and Steven Shavell, Punitive Damages: An Economic Analysis, 111 Harv. L. Rev. 869 (1998); Marc Galanter and David Luban, Poetic Justice, 42 Am U L Rev 1393 (1993); David G. Owen, The Moral Foundations of Punitive Damages, 40 Ala. L. Rev. 705 (1989); William Landes and Richard A. Posner, The Economic Structure of Tort Law 160-65, 184-85, 223-24 (1987)..

¹⁶ See The New York Times, July 31, 1999, Section A; Page 9; Column 3, General Motors Appeals Record Lawsuit Damages.

¹⁷ See *BMW v. Gore*, 116 S. Ct. 1589 (1996).

¹⁸ See generally Developments in the Law—The Civil Jury, 110 Harv L Rev 1408, 1533 n. 158 (1997) (discussing actual and proposed caps).

¹⁹ See, e.g., Product Liability Reform Act of 1997, S. 5, 105th Cong. Section 108 (1997). For general discussion, skeptical of the attention paid to high awards, see Marc Galanter, *Shadow Play: The Fabled Menace of Punitive Damages*, 1998 Wisconsin L Rev 1.

²⁰ See Judy Shih and Cass R., Sunstein, Damages in Sexual Harassment Cases, in *Sexual Harrassment* (Catharine MacKinnon and Reva Siegel eds, forthcoming 2000).

for example, compensatory damages for pain and suffering, libel, and intentional infliction of emotional distress.²¹ Similar problems arise whenever an administrative agency is asked to impose civil fines and also in the area of criminal sentencing.²²

Participants in the legal system are often requested to come up with some kind of judgment, factual or normative, and then to “translate” that judgment into a dollar award. In the area of punitive damages, it is necessary to make some assessment of the character of the defendant’s behavior, and then to ascertain the appropriate dollar amount to be paid to the plaintiff by way of punishment. In many domains, compensatory judgments raise similar puzzles. While juries are nominally expected to find a “fact”—what amount of money would restore the plaintiff to the status quo ante?—it is often extremely difficult to monetize the relevant harm, and normative judgments undoubtedly play a significant role.²³ In the case of punitive damages, it is extremely difficult for even experts to agree on what dollar amount constitutes adequate “punishment” or produces an appropriate deterrent signal.

In all of these areas, the legal system is pervaded by a degree of unpredictability and variance, resulting in apparent arbitrariness, as similarly situated people are treated differently.²⁴ An extensive study of pain and suffering cases found that as much as 40% of the awards consists of “noise,” unexplained by objective factors.²⁵ A study of all reported sexual harassment cases was unable to connect either compensatory or punitive awards to any case characteristics that

²¹ See, e.g., Randall Bovgher et al., Valuing Life and Limb in Tort, 83 NW U L Rev 908 (1989); David Baldus et al., Improving Judicial Oversight of Jury Damages Assessments, 80 Iowa L Rev 1109 (1995); David Leebron, *supra* note.

²² See Edward L. Rubin, Punitive Damages, 1998 Wisc. L. Rev. 131.

²³ See generally Michelle Anderson and Robert MacCoun, Goal Conflict in Juror Assessments of Compensatory and Punitive Damages, 23 Law and Human Behavior 313 (1999); Patrick Atiyah, The Damages Lottery (1997).

²⁴ See, e.g., David Leebron, *supra* note; Jonathan Karpoff and John Lott, On the Determinants and Importance of Punitive Damages Awards, 42 J. L. & Econ. 527 (1998).

²⁵ See Leebron, *supra* note.

might be thought to explain jury judgments.²⁶ The punitive damage area is more complicated—a point to which we will return shortly—but there is evidence of significant variability here as well.²⁷ The most ambitious claims to the contrary attempt to show that once the compensatory award has been made, the punitive award becomes predictable to a certain degree²⁸; but the same data show that at the time a case is filed, it is very hard to know the expected punitive award, and that there is a generally a great deal of “noise” in outcomes.²⁹

To understand the current study, it is necessary to understand its predecessor by the way of background. Our earlier study involved a demographically diverse set of jury-eligible citizens from Austin, Texas.³⁰ The relevant experiment involved 28 personal injury cases, which respondents were asked to assess in one of three ways: outrageousness, on a bounded numerical scale (0 to 6); intent to punish, on a bounded numerical scale (also 0 to 6); and actual

²⁶ See Judy Shih and Cass R. Sunstein, *Damages in Sexual Harassment Cases*, forthcoming in *Sexual Harassment* (Catharine MacKinnon and Reva Seigel eds. 1999)

²⁷ See Karpoff and Lott, *supra* note. There is some dispute over the degree of unpredictability. Theodore Eisenberg et al., *The Predictability of Punitive Damages*, 26 *J. Legal Stud.* 623 (1997), shows that the log of punitive awards is predicted reasonably well from a set of objective characteristics of cases in which awards were made; in particular, it is shown that the compensatory award is a fairly good predictor of the punitive award. See *id.* at 644. In terms of real dollars rather than log dollars, however, there is a degree of unpredictability in the Eisenberg data as well, because the severe skewness of the awards creates a possibility of either small or huge awards in identical cases. See also Karpoff and Lott, *supra*.

Note also that predictability can be understood in different ways: (a) predictability exists when case characteristics predict punitive awards; (b) predictability exists when the judgments of one group of six or twelve predicts the judgments of another group of six or twelve; (c) predictability exists when an actor can assess expected liability when something goes wrong. Our principal emphasis here is on (b); Eisenberg's emphasis is on (a); both are relevant to (c). Of course the three are closely related in practice.

²⁸ See Theodore Eisenberg et al., *The Predictability of Punitive Damages*, 26 *J. Legal Stud.* 623 (1997).

²⁹ See Karpoff and Lott, *supra* note.

³⁰ See Cass R. Sunstein, Daniel Kahneman, and David Schkade, *Assessing Punitive Damages*, 107 *Yale LJ* 2071 (1998).

awards, on the unbounded scale of dollars. As noted, our principal findings were twofold. People's moral judgments are widely shared and predictable, in fact strikingly so, at least in the personal injury cases studied in the experiment.³¹ But in spite of this point, and in the presence of shared moral judgments, people's judgments on a dollar scale—the scale, or “response mode,” favored by the legal system—are highly unpredictable in the sense that the median judgment of any group of twelve people is an extremely poor predictor of the median judgment of any other group of twelve people.³² Lacking a reliable understanding of how deliberation would affect individual judgments, we used the median of groups of twelve individuals, randomly selected from our pool of 899 and combined into a large number of “synthetic juries.”³³ At least in the experimental setting, the primary identifiable source of the noise is the difficulty jurors have in translating their punitive intent into dollars. Dollar awards are highly variable despite the existence of shared moral judgments.

To explain why the use of the dollar scale would produce variability, we explored a link between the dollar scale and psychological research on the problem of “magnitude scaling,” which occurs when people are asked to assess stimuli—the brightness of lights, the loudness of noise—along an unbounded numerical scale.³⁴ This research shows a great deal of variability in assessments. The underlying problem is that people are being asked to scale without a “modulus,” that is, without a standard that would help give meaning to various numbers on the scale.³⁵ Consider the words of one of the subjects in the relevant experiments: “I felt freer to use numbers over a wide range. I liked the idea that I could just relax and contemplate the tones. When there was a fixed standard I felt more constrained

³¹ Id. at 2097-2100.

³² See id. at 2100-2103.

³³ We relied on evidence that median judgments are a good predictor. See James Davis, Group Decision Making and Quantitative Judgments: A Consensus Model, in *Understanding Group Behavior* 35, 47 (Erich White and James Davis eds., 1996); Shari Diamond and Jonathan Casper, *Blindfolding the Jury*, 26 *L. & Society Rev.* 513, 553 (1992). We did note, however, the possibility of effects of the sort observed in the study here. See SKS at 2101 at n. 128.

³⁴ See S.S. Stevens, *Psychophysics* (1975).

³⁵ SKS at 2106-2107.

to try to multiply and divide loudnesses, which is hard to do; but with no standard I could just place the tone where it seemed to belong.”³⁶

The key point is that when a modulus is supplied, the variability greatly decreases. Juries asked to assess punitive damage awards are in effect asked to scale without a modulus. Unpredictable judgments are a natural result even when people do not disagree about the significant issues³⁷—even when, that is, there is a kind of “bedrock” moral judgment in place. If this point is correct, it helps explain the observed variability in dollar awards in many areas of the law. It also helps explain the disparities that led to the enactment of the Sentencing Guidelines³⁸; before the guidelines, judges were being asked, in effect, to scale without a modulus, since the relevant scale (years) has a great deal in common with the dollar scale (i.e., bounded below at zero, but with great discretion at the high end).

Our earlier study did not, however, involve deliberating juries, and a natural question is whether deliberating juries would produce similar or quite different results. Perhaps deliberation would reduce variability.³⁹ In any case a test of deliberating juries would help to

³⁶ See S.S. Stevens, *Psychophysics* 37 (1975).

³⁷ For similar results, see Michael J. Saks et al., *Reducing Variability in Civil Jury Awards*, *Law & Human Behav.*, Spring 1997, at 243.

³⁸ See Sandra Shane-DuBow et al., *US Department of Justice, Sentencing Reform in the United States* (1985).

³⁹ In a series of papers, Michael Saks has argued that juries actually reduce variance. Saks’ research began by comparing twelve person juries to six person juries, see the overview in Michael Saks, *The Smaller the Jury, the Greater the Unpredictability*, 79 *Judicature* 263 (1996)—a comparison on which our study here does not bear. But the analysis has been extended to comparisons of juries and judges, with the suggestion that juries are likely to produce less variability by virtue of their numbers. See Michael Saks, *Do We Really Know Anything About the Behavior of the Tort Litigation System—and Why Not?* 140 *University of Pennsylvania Law Review* 1147 (1992); *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, with Peter D. Blanck, 44 *Stanford Law Review* 815 (1992); *Reducing Variability in Civil Jury Awards*, with Lisa A. Hollinger, Roselle L. Wissler, David L. Evans, and Allen J. Hart, 21 *Law and Human Behavior* 243 (1997). This basic conclusion is briefly challenged in a footnote in Robert MacCoun, *Inside the black box: What empirical research tells us about decisionmaking by civil juries*, in *Verdict: Assessing the civil jury system* 137, 178 n. 26 (Robert E. Litan ed. 1993): “The argument is based on statistical sampling theory, but

confirm or deny the wisdom of the decision, for purposes of creating synthetic juries, to treat the median judgment of a group of twelve as the likely judgment of any deliberating group. An alternative possibility, referred to in the introduction, was to assume that there would be group polarization—that the process of collective deliberation would move the group further in the direction of the initial tendency suggested by the median of individual judgments. Hence our main purpose in this study was to examine the effects of jury deliberation on dollar awards and in particular to see whether deliberation would increase or decrease predictability. In the process we also hoped, as a secondary goal, to find out whether the original findings—strikingly shared moral judgments but erratic awards—would be replicated with a new sample of citizens from a different state, and with new and richer case materials.

III. Deliberating Juries: An Experimental Inquiry

A. Method

Jury-eligible citizens from Phoenix, Arizona were recruited and paid by a survey firm. Participants were randomly assigned to a six-person jury, and to a response mode order; some juries judged dollar awards first and punishment ratings second, and others completed the tasks in the opposite order. Six juries (out of a total of 480) had only five members because an insufficient number of participants showed up at a given appointment time. A pilot test of 29 juries was conducted in Phoenix to test the materials and procedure. Because adjustments were very minor, these juries were added to the main sample and the combined sample was analyzed together. Therefore, a total of 3048 citizens participated in 509 juries.

the analogy between empaneled juries and random samples is an imperfect one. Thus, it is a plausible hypothesis, but requires more rigorous testing than it has received to date.” We have attempted a more rigorous test here, finding that juries produce more variability as compared with the mean of individual predeliberation judgments. This finding suggests, though it does not prove, that juries will produce more variability in awards than judges (a suggestion supported by the possibility that judicial experience with a wide range of cases will introduce the equivalent of a “modulus” by which to discipline dollar awards).

The procedure consisted of four parts. In Part 1, all participants in a given session viewed a videotape for the case they would consider, read the corresponding written materials, and recorded their personal judgment of the appropriate punitive damage award or punishment rating (Table 1). In Part 2, participants were randomly assigned to a jury of six members, and given 30 minutes to deliberate on and reach a unanimous verdict on a punitive damage amount or a severity of punishment rating.

Table 1
Response Mode Manipulation

Punishment

How much should the defendant be punished because of their actions, and to deter the defendant and others from similar actions in the future? Note that the compensatory damages that the defendant must pay do not count as part of the punishment. Please circle the number that best expresses the *jury's* judgment of the *appropriate level of punishment*.

None	Mild	Substantial	Severe	Extremely Severe				
0	1	2	3	4	5	6	7	8

\$ 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? Note that the compensatory damages that the defendant must pay do not count as part of the punishment. Please write the *amount of punitive damages* that the *jury* agreed on in the blank below.

\$ _____

In Part 3, a new individual response form was distributed, which asked them to record a second personal judgment for the

same case, using the complementary type of verdict (punishment rating or dollar damages) from the one they had already used. In Part 4, the jury again deliberated to reach a unanimous verdict on this second type of judgment for the same case. Thus, for each individual, and for each jury, we have both a dollar award and a punishment rating for the case they considered.

The stimuli consisted of fifteen personal injury scenarios.⁴⁰ An example is provided in the Appendix. A videotape was prepared for each case, in which a professional actor read the text of the case and all instructions aloud. To maximize comprehension, participants were required both to view the videotape and to read the written version. Firm size (annual profits of \$100-200 million) and compensatory damages (\$200,000) were held constant for all cases. Thus, the variability we observe cannot be accounted for by a model that depends on variability in compensatory damage awards.

B. Results

1. Basics

Despite the half hour time limit for deliberation, 91% of juries reached a unanimous verdict on the punishment scale (a total of 461 verdicts) and 82% of juries reached a unanimous verdict on a dollar amount (a total of 416 verdicts). The remainder had not reached a verdict when the time limit expired; these were treated as hung. All further analyses were conducted on the 401 juries that reached *both* a punishment verdict and a dollar verdict.⁴¹ Because there were no significant differences between response mode orders (i.e., dollars-punishment or punishment-dollars), we pooled verdicts made by dollar-first juries and dollar-second juries in our analyses (punishment verdicts were treated similarly).

⁴⁰ Of these ten were more elaborate versions of the same scenarios used in Kahneman, Schkade and Sunstein. *supra*, and five were completely new scenarios which, like the first 10, were based on real cases (Table 1). The main substantive elaboration on the original scenarios was the addition of a paragraph of closing arguments by attorneys for each side.

⁴¹ We chose the more conservative path of focusing on juries with complete responses to ensure that comparisons between punishment and dollar verdicts, and between individuals and juries, were based on the same set of respondents. Recreating our Tables and Figures with all available responses produces the same pattern of results, with some slight differences in exact numbers.

Table 2
Summary of Personal Injury Scenarios

<i>Case</i>	<i>Description</i>
Williams v. National Motors	Motorcycle driver injured when brakes fail
Smith v. Public Entertainment	Circus patron shot in arm by drunk security guard
Douglas v. Coastal Industries	Auto airbag opens unexpectedly, injuring driver
Sanders v. A&G Cosmetics	Man suffers skin damage from using baldness cure
Stanley v. Gersten Productions	Elderly woman suffers back injuries from using exercise video
Glover v. General Assistance	Child ingests large quantity of allergy medicine, needs hospital stay
Lawson v. TGI International	Employee suffers anemia due to benzene exposure on the job
Newton v. Novel Clothing	Small child playing with matches burned when pajamas catch fire
West v. MedTech	Disabled man injured when wheelchair lift malfunctions
Windsor v. Int. Computers	Secretary chronically ill due to radiation from computer monitor
Reynolds v. Marine Sulphur	Seaman injured when molten sulphur container fails
Crandall v. C&S Railroad	Train hits car at crossing, injuring driver
Dulworth v. Global Elevator	Shopper injured in fall when escalator suddenly stops
Hughes v. Jardel	Store employee raped in mall parking lot
Nelson v. Trojan Yachts	Man nearly drowns when defective boat sinks

In the aggregate, jury punishment and dollar verdicts produced very similar rankings of the cases: there is a rank correlation of .88 between the jury dollar and punishment verdicts in Table 3.⁴²

⁴² The Spearman rank correlation is an index of agreement between rankings that is analogous to first converting each column to ranks (from 1 to 15 in this

Table 3
Aggregate Responses by Case and Condition

<i>Case</i>	<i>Individuals</i>		<i>Juries</i>	
	<i>Mean Punishment Rating</i>	<i>Median \$ Damage Award</i>	<i>Mean Punishment Rating</i>	<i>Median \$ Damage Award</i>
Reynolds	5.52	1,500,000	5.83	10,000,000
Glover	5.15	1,000,000	5.28	4,000,000
Lawson	4.42	500,000	4.64	2,000,000
Williams	4.82	500,000	5.27	1,500,000
Smith	5.372	50,000	5.75	1,000,000
West	4.54	500,000	5.42	1,000,000
Nelson	4.86	500,000	5.22	1,000,000
Hughes	4.83	500,000	5.14	1,000,000
Crandall	4.05	200,000	4.18	500,000
Douglas	3.77	200,000	4.07	500,000
Sanders	2.71	75,000	2.48	100,000
Windsor	2.56	50,000	1.93	50,000
Stanley	1.48	0	1.38	0
Dulworth	1.36	0	1.13	0
Newton	1.10	0	0.77	0
Average of Top 5	5.06	750,000	5.36	3,700,000
Average of Middle 5	4.41	380,000	4.81	800,000
Average of Bottom 5	1.84	25,000	1.54	30,000
Overall Average	3.77	385,000	3.90	1,510,000

case) and then computing the correlation between the two sets of ranks. It is interpreted similarly to conventional correlations.

2. High awards way up, low awards slightly down

A view of the overall results for individuals and juries in Table 3 appears to show a striking and simple picture: juries produce higher dollar awards than the individuals who compose them. This pattern is also reflected, to a lesser degree, in punishment ratings. Higher jury awards are not confined to the middle of the award distribution, and also show up clearly in the extreme low (10th percentile) and high (90th percentile) ranges of the distribution as well (Table 4).

Table 4
Dollar Punitive Damage Distributions
by Case for Individuals and Juries (10/90)

<i>Case</i>	<i>Individuals</i>			<i>Juries</i>		
	<i>10th pctile</i>	<i>Median</i>	<i>90th pctile</i>	<i>10th pctile</i>	<i>Median</i>	<i>90th pctile</i>
Reynolds	150,000	1,500,000	15,000,000	1,000,000	10,000,000	50,000,000
Glover	50,000	1,000,000	10,000,000	1,000,000	4,000,000	50,000,000
Lawson	0	500,000	10,000,000	250,000	2,000,000	15,000,000
Williams	25,000	500,000	5,000,000	200,000	1,500,000	10,000,000
Smith	1,000	250,000	5,000,000	100,000	1,000,000	20,000,000
West	2,000	500,000	3,000,000	250,000	1,000,000	5,000,000
Nelson	25,000	500,000	5,000,000	200,000	1,000,000	20,000,000
Hughes	30,000	500,000	3,250,000	200,000	1,000,000	4,000,000
Crandall	0	200,000	2,000,000	50,000	500,000	2,000,000
Douglas	0	200,000	3,250,000	1	500,000	25,000,000
Sanders	0	75,000	1,000,000	0	100,000	1,000,000
Windsor	0	50,000	1,000,000	0	50,000	5,000,000
Dulworth	0	0	200,000	0	0	200,000
Newton	0	0	500,000	0	0	300,000
Stanley	0	0	200,000	0	0	250,000
Medians	0	250,000	3,250,000	100,000	1,000,000	5,000,000

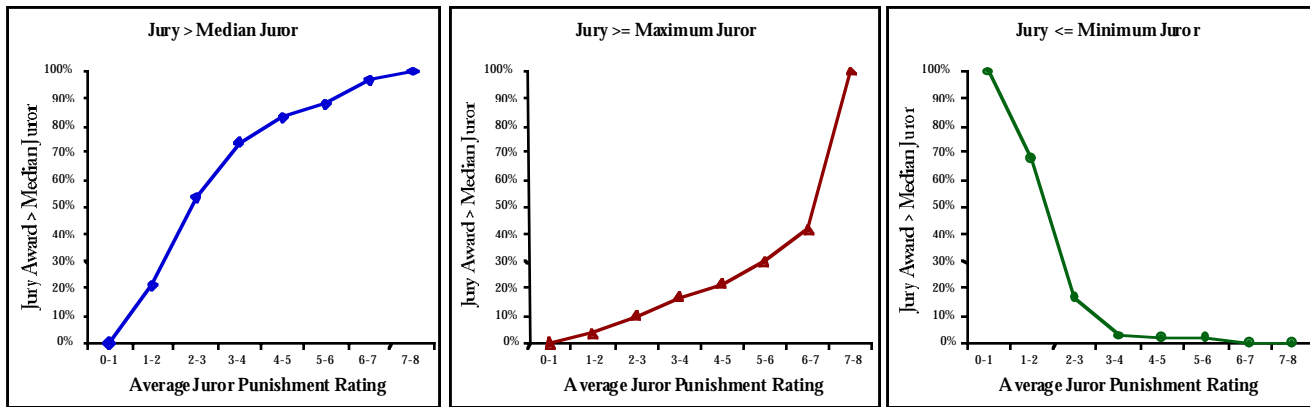
This dramatic upward shift is indeed an important part of the picture, but it does not tell the whole story. A closer look at Table 3 reveals a systematic difference among cases in the degree to which juries are harsher than individuals: the more egregious the case, the greater is the margin by which jury awards exceed individual awards. For the five cases with the lowest awards, individual and jury awards are approximately *equal*, but for the middle five cases jury awards are *twice* as large, and for the top five cases *five* times as large. Indeed, on the punishment scale, deliberation actually causes cases the bottom five cases (those below the midpoint of the scale) to *drop*, and those at or above the midpoint to *rise*.

As we will see, this is a characteristic response pattern when group polarization is present. Thus it appears that polarization occurs in the underlying moral judgments, as indicated by the punishment ratings, and that the effect of this shift is amplified when expressions must be made on the unbounded dollar scale. The fact that dollar awards for the least severe cases did not decline is probably a consequence of the absolute minimum award of zero, below which awards could not go, rather than of a lack of a shift in underlying judgments.

A more direct measure of polarization can be computed by comparing jury verdicts to the judgments of the specific jurors that compose them, regardless of which case they considered. If group polarization is indeed occurring, we would expect that juries whose individuals initially judge a case to be more severe would more frequently produce high jury verdicts, compared to the predeliberation judgments of individuals. The reverse would be true of juries whose individuals initially judged a case to be less severe. Note that this contrast can happen for the same legal case, depending on the individual judgments jurors bring to deliberation.

To summarize the pre-deliberation judgments of jurors, we use three logical benchmarks as points of comparison: the median, the maximum, and the minimum judgments of the individuals in a jury. A clear pattern emerges on all three of these measures, as depicted in Figure 1. As a jury's individuals judge a defendant's behavior to deserve more severe punishment, the tendency of the jury to exceed the median dollar award of its members rises rapidly, reaching 100% at an average rating of 7 or more.

Figure 1
 Juror Judgments Compared to the \$ Awards of Their Juries



The tendency toward extremely high verdicts shows a similar pattern: the percentage of jury awards at or above the maximum individual award rises steadily with the intended severity of punishment, and indeed leaps to 100% when the average juror is at or above 7 on the punishment scale. In contrast, the tendency toward extremely low verdicts shows the opposite pattern: the percentage of jury awards at or below the minimum individual award declines with the intended severity of punishment, and leaps to 100% when the average juror is below 1 on the punishment scale.

As a result of the increased extremity produced by polarization, juries produce more unpredictable judgments than individuals, in the sense that the width of a confidence interval for a prospective award is much greater.⁴³ Jury awards also have a higher standard deviation ($\sigma = \$14,118,179$) than do the median awards of their jurors ($\sigma = \$2,715,098$). Contrary to the possibility that this increased variance results from juries distinguishing more consistently between cases than individuals, this relationship holds separately for each of the 15 cases. The problem is less severe but still present with jury punishment ratings ($\sigma = 2.37$) compared to individuals ($\sigma = 2.00$).⁴⁴

A important final question about the increased unpredictability is the extent to which it results from additional random noise introduced by the uncontrolled vagaries of group deliberation, or instead from the systematically higher level of awards due to polarization. Because dollars are a magnitude scale, which is bounded at zero, a higher average award necessarily means a higher variance as well. Indeed, a well understood characteristic of magnitude scales is a high correlation between the mean and standard deviation of responses. To investigate the source of the higher variance, we borrow a classic analysis from psychophysics research, which (for those who wish to skip the technical details) establishes that the increased variance is indeed due primarily to a higher mean and relatively little to random deliberation noise. If the shift in the level of awards causes most of the increased variance, then the

⁴³ For example, in Table 4 an 80% confidence interval for a jury award for the Glover case would be \$1,000,000 to \$50,000,000 and an 80% confidence interval for an individual award would be \$50,000 to \$10,000,000.

⁴⁴ Thus the statistical juries analyzed in our previous study substantially understated the unpredictability of jury awards.

relationship between the mean and the standard deviation of awards for different cases should be the same for juries and their individuals. As is expected of a magnitude scale, the relationship is strong for both individuals ($R^2 = .66$) and juries ($R^2 = .79$) and, remarkably, both sets of points fall on essentially the same regression line.⁴⁵ This strongly suggests that the mechanism by which jury awards are less predictable is a product not of irrational or erratic deliberations, but rather of higher average awards, resulting from the more orderly and predictable process of polarization.

3. Predictability

How predictable are our jury judgments? One sense in which jury judgments can be considered predictable is if there is high agreement between juries randomly selected from the population. To examine this question, we reprise an analysis from our earlier study, which simulated the convening of a separate jury for each case on a given day. By comparing the verdicts of different sets of randomly selected juries for our 15 cases we can assess the predictability of awards.⁴⁶ The correlations in Table 5 show that there is strong agreement between sets of juries when judging punishment ($r = .69$), but far less agreement when they judge dollar awards ($r = .24$).⁴⁷ This correlation provides a striking conclusion: only 6% of the variance⁴⁸ in dollar awards is due to differences between cases, and the remaining 94% is random noise. In contrast,

⁴⁵ The slopes of the two regression lines are not significantly different.

⁴⁶ The basic unit of this analysis is a set of 15 jury judgments (one for each of the 15 cases), where one of the jury judgments for a given case is randomly selected. This procedure simulates the convening of independent juries to deal with 15 separate cases on the same day. Using this procedure, we created 60 sets of 15 jury judgments for each response mode. We then computed the Pearson correlation between each pair of sets. This computation was performed both within response (e.g., the correlation between the punishment ratings of different sets of 15 juries) and across responses (e.g., the correlation between the punishment rating of set of 15 juries and the dollar awards of another set). The data shown in Table 5 are medians of the 1770 correlations obtained within each response mode or of the 3600 correlations obtained between two response modes.

⁴⁷ Further, the correlation between different sets of 15 juries responding with dollar awards is actually lower than the correlation between a set of juries judging punishment and another judging dollars.

⁴⁸ The percentage of variance explained is equal to the square of the correlation.

almost 50% of jury punishment ratings is accounted for by differences between cases which, while less than perfect, offers a dramatic improvement over dollar awards.

Table 5
Correlations Between Judgments of Sets of Mock Juries

	Punishment	Dollars
Punishment	.69	
Dollars	.33	.24

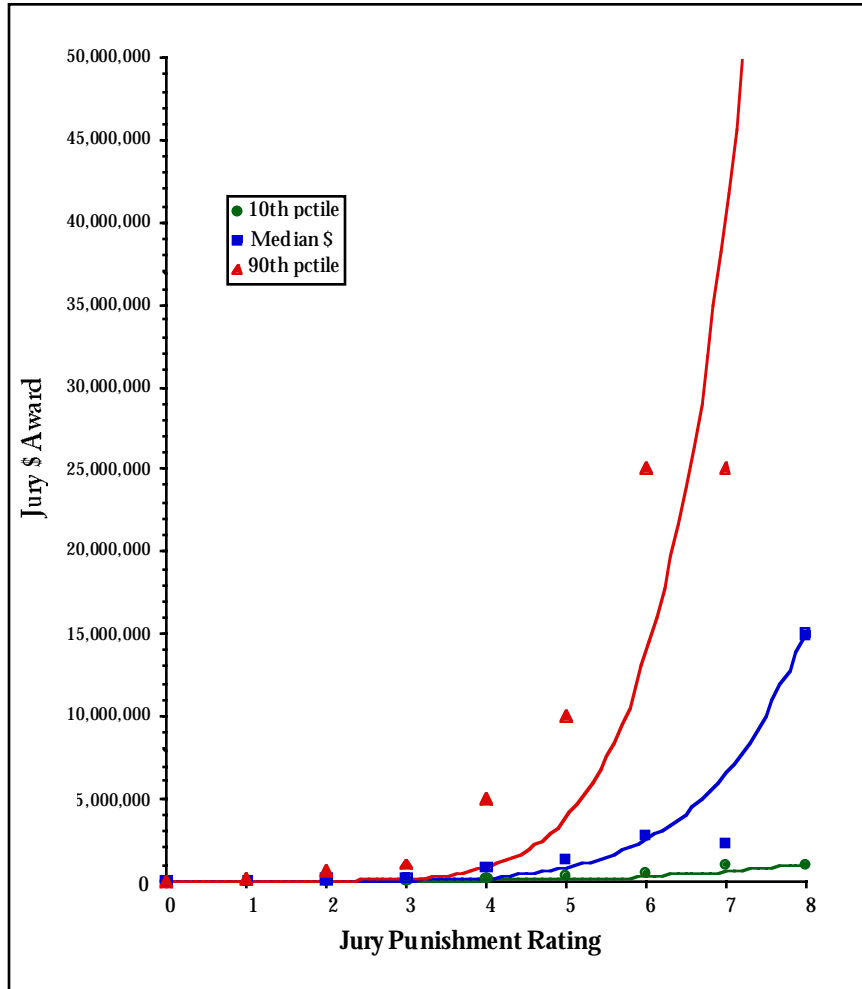
All entries are average correlations between 60 vectors of 15 randomly selected jury awards (one for each case).

Table 6
Distribution of Jury \$ Awards for
a Given Jury Punishment Rating

Jury Punishment Rating	Jury Dollar Awards				
	10th pctile	Median\$	90th pctile	Mean \$	Juries
8	1,000,000	15,000,000	100,000,000	32,583,333	12
7	1,000,000	2,250,000	25,000,000	10,205,000	20
6	500,000	2,750,000	25,000,000	9,756,757	74
5	250,000	1,333,333	10,000,000	4,857,111	75
4	200,000	825,000	5,000,000	3,436,947	90
3	40,000	200,000	1,000,000	595,000	33
2	0	100,000	600,000	399,842	19
1	0	22,500	100,000	40,063	16
0	0	0	0	239	63

Another sense of predictability, one of great practical importance, concerns the relationship between a jury's underlying moral judgment and its dollar award. If juries who agree perfectly on

Figure 2
 The Greater the Intent to Punish,
 the Greater the Range of Possible Awards



punitive intent disagree on a dollar award, the similarly situated are plainly treated differently. To examine this form of unpredictability, we analyze the distribution of dollar awards for juries who agreed on exactly the same punishment rating (Table 6). It is clear from Table

6 that even juries who agree precisely on punitive intent vary widely in the dollar awards they judge to be appropriate. For example, at the most common punishment rating of 4 (“Substantial” punishment), awards range from \$200,000 at the low end (the 10th percentile) to \$5,000,000 at the high end (the 90th percentile), and the maximum award (not listed) is dramatically higher at \$100,000,000. Note that the range between the 10th and 90th percentiles for particular punishment ratings in Table 6 are quite comparable to those for individual cases in Table 4. As we found for those different cases, the greater the intent to punish, the more extensive is the range of possible awards, as depicted in Figure 2. The mapping of punitive intent onto the magnitude scale of dollars is clearly implicated as the source of this aspect of unpredictability, since the moral judgment, as well as the compensatory damages and firm size, are all held constant.

4. Do People From Arizona Agree With People From Texas? The Effects of Geography, Race, Gender, Education, Age, and Wealth

A subsidiary but nonetheless important question is whether the findings of the earlier study are replicated under the current study’s changes in stimuli, procedure, and sample. The answer is that the previous results are replicated in every essential respect. The findings in the Texas study were replicated in Arizona, despite evident differences between the two regions, and people from the two areas evaluated cases in the same way. As before, dollars and ratings produce very similar rankings of the cases (a rank correlation of .90 compared to .91 in the previous study). Different demographic groups again produced very similar average evaluations, as indicated by the extremely high correlations in Table 7. Punishment ratings in Table 7 also decisively outperformed untransformed dollar awards in distinguishing between cases (Table 8), in the sense that differences between cases accounted for only 1% of the variance in individual dollar awards (vs 34% for punishment ratings) and only 10% of jury dollar awards (vs 53%).

Table 7
Correlation Between Demographic Groups
on Intended Severity of Punishment^a

Gender		<i>Men</i>		
	Women	.99		
Ethnicity		<i>White</i>	<i>Hispanic</i>	
	Hispanic	.92		
	Other	.88	.81	
Household Income		<i>< 30K</i>	<i>30-50K</i>	
	30-50K	.98		
	> 50K	.99	.99	
Age		<i>< 30</i>	<i>30-39</i>	<i>40-49</i>
	30-39	.97		
	40-49	.96	.97	
	> 50	.96	.97	.97

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

Table 8
 Proportion of Variance Explained by Cases:
 Comparison to Kahneman, Schkade & Sunstein (1998)

	<i>KSS98*</i>	<i>Individuals</i>	<i>Juries</i>
Punishment ratings	.40	.34	.53
Raw \$ awards	.03	.01	.10
Log \$ awards	.46	.31	.54

* KSS98 numbers are from the large firm/isolation condition, which is directly comparable to the individual condition in the current study.

In addition, the ordering of case evaluations closely matches that in our previous study. There are ten cases that are common to both studies, and evaluations made by Texans in the previous study are highly predictive of those made by Arizonans in the current study—the rank correlation between the two samples is .90 for punishment ratings and .98 for dollar awards. Finally, the absence of an effect of response mode order suggest that prior deliberation about a case, and even agreeing on a punishment rating, does little to reduce unpredictability in dollar awards. Thus, the current larger study, with several nontrivial changes, confirms the conclusion of our previous study that individual moral judgments are predictable and shared, but that expressing them in dollars produces unpredictability and confusion, and especially so in juries.

IV. What Happened?

Deliberation and Group Polarization

Plainly many of the 509 deliberating juries were subject to group polarization. Indeed, we believe that this is the first experimental evidence of group polarization in punitive damage awards.⁴⁹ It is highly likely that the same pattern would be found

⁴⁹ Group polarization has been shown with simulated juries in the context of guilt and innocence. See David Myers and Martin Kaplan, Group-Induced Polarization in Simulated Juries, 2 *Personality and Soc. Psych. Bulletin* 63 (1976).

with compensatory awards involving hard-to-monetize injuries, such as libel, civil rights violations, pain and suffering, and emotional distress. In this section, we provide a general discussion of group polarization, both to explain our results here and because the phenomenon, which has received no sustained attention in the legal literature, is of general interest.

A. The Basic Phenomenon

Group polarization is one of the most common patterns of difference between individual decisions and deliberating groups,⁵⁰ and it has been found in many diverse tasks. Polarization “is said to occur when an initial tendency of individual group members toward a given direction is enhanced [by] group discussion.”⁵¹ The result is

It might be tempting to suggest that the results are partly or largely a function of the decision rule, in this case unanimity. Perhaps the requirement of unanimity pushed people further in the direction of the dominant view, without any group polarization, an idea that might be fortified with the thought that those with outlier positions (in favor of extreme awards) would be especially likely to hold out against a compromise view, thus producing pressure toward the extremes. In the abstract it seems plausible to suggest that the decision rule would matter, but as an explanation for our findings, it seems less plausible in light of the fact that 20% of the jury judgments were as high as or higher than that of the highest individual judgment pre-deliberation. Note also that numerous studies show that group polarization occurs regardless of the decision rule and hence it is extremely unlikely that the unanimity rule accounted for other results here: “The shift effect is about equally robust regardless of whether a group decision is required.” See David Myers and Helmut Lamm, *The Group Polarization Phenomenon*, 83 *Psych. Bulletin* 602, 611 (1976). Of course we cannot exclude the possibility that the results would be somewhat different without a unanimity rule; this is in fact a good area for subsequent empirical study, especially in light of continuing questions about the consequences of requirements of jury unanimity.

⁵⁰ Some studies suggest that the median predeliberation judgment is a good predictor of outcomes; but in some situations groups have indeed been observed to make quite different decisions from those of the median or average of individuals that compose them. See Norbert Kerr, Robert MacCoun, and Geoffrey Kramer, *Bias in Judgment: Comparing Individuals and Groups*, 103 *Psychological Review* 687 (1996); Daniel Gigone and Reid Hastie, *Proper Analysis Of The Accuracy Of Group Judgments*, 121 *Psychological Bulletin* 149 (1997).

⁵¹ See Isenberg, *supra* note, at 1141.

that groups often make more extreme decisions—both higher or lower—than would the typical or average individual in the group.

Consider some examples. (a) A group of moderately profeminist women will be more strongly profeminist after discussion.⁵² (b) Citizens of France become more critical of the United States and its intentions with respect to economic aid.⁵³ (c) After discussion, whites predisposed to show racial prejudice offer more negative responses to the question whether white racism is responsible for conditions faced by African-Americans in American cities.⁵⁴ (d) After discussion, whites predisposed not to show racial prejudice offer more positive responses to the same question.⁵⁵ It should follow, for example, that a group moderately predisposed in favor of affirmative action should strongly favor it after discussion; that those moderately critical of an ongoing war effort would, after discussion, sharply oppose the war; that a group moderately predisposed to hire a certain job candidate will, after discussion, support the application with considerable enthusiasm. The phenomenon has conspicuous importance to the operation of deliberating bodies, including juries; but it has been barely noticed in the legal culture. Obviously the pattern described above is exactly what would be predicted if group polarization were at work.⁵⁶

⁵² See D.G. Myers, Discussion-Induced Attitude Polarization, 28 *Human Relations* 699 (1975).

⁵³ Brown at 224.

⁵⁴ D.G. Myers and G.D. Bishop, The Enhancement of Dominant Attitudes in Group Discussion, 20 *J Personality and Soc. Psych.* 286 (1976),

⁵⁵ See *id.*

⁵⁶ There is one difference: In the group polarization studies, the phenomenon is usually *defined* by reference to scales having two sides, with a “neutral” midpoint. This is the arrangement by which it makes sense to speak of initial dispositions and their aggravation. Dollar awards, by contrast, have no “neutral” midpoint. Our study suggests that what matters is the psychological predisposition—to increase the tendency in which the group is leaning—rather than the methodological definition. Hence the processes that produce polarization occur even in the absence of a defined midpoint. We are grateful to Robert MacCoun for help with this point.

B. Risky Shifts, Cautious Shifts

Group polarization was first found in a series of experiments involving risk-taking decisions.⁵⁷ Before 1961, conventional wisdom had been that as compared with the individuals who compose it, a group of decision-makers—for example a committee or board—would be likely to favor a compromise and thus to avoid risks. But the relevant experiments found otherwise; they identified what has become known as the “risky shift.”⁵⁸ A group of deliberating people was more likely to take risks than are the individual members.

In these experiments, people were asked to assess the probability that a new company, for which some person A is considering employment, will prove “financially sound.” Group decisions showed a repeated pattern toward greater risk-taking—that is, after discussion, participants tended to assess the likelihood of financial soundness as consistently higher than the median judgment of the group predeliberation. But later studies showed that under certain conditions, it was possible to induce a “cautious shift” as well. Where the judgments of individual group members were relatively cautious, deliberation would produce a strong tendency toward greater caution. Thus “a group discussion moves decisions to more extreme points in the direction of the original inclination . . . , which means shift to ether risk or caution in the direction of the original disposition, and the size of the shift increases with the degree of the initial polarization.”⁵⁹ Similar results have been found in many contexts, involving, for example, questions about economic aid, architecture, political leaders, race, feminism, and judgments of guilt or innocence.⁶⁰ The parallel here is of course the increase in large awards (for juries whose members were individually disposed in this direction) and the decrease in small awards (for juries whose members were originally so disposed).

⁵⁷ We draw in this and the following paragraph on Brown, *supra* note, at 200-206.

⁵⁸ See J.A.F. Stoner, *A Comparison of Individual and Group Decisions Including Risk*, unpublished master’s thesis, School of Management, Massachusetts Institute of Technology; J.A.F. Stoner, *Risky and Cautious Shifts in Group Decisions*, 4 *J Experimental Social Psych.* 442 (1968).

⁵⁹ Brown, *supra*, at 211.

⁶⁰ See *id.*

C. Two Mechanisms

There have been two main explanations for group polarization, both of which have been extensively investigated.⁶¹ Massive support has been found on behalf of both explanations.⁶²

The first involves *social comparison*.⁶³ On this view, people want to be perceived favorably by other group members (and also to perceive themselves favorably), and once they hear what others believe, they adjust their positions in the direction of the dominant position. They may want to signal, for example, that they are not cowardly or cautious, and hence they will frame their position so that they do not appear such by comparison to other group members.⁶⁴ The result is to press the group's position toward one or another extreme.

The dynamic behind the social comparison explanation is that most people may want to take a position of a certain socially preferred sort—in the case of risk-taking, for example, they may want to be perceived (and to perceive themselves) as moderate risk-takers, and their choice of position may be partly a product of this

⁶¹ Isenberg, *supra*, and Brown, *supra*, review this literature. We draw largely on work done in the 1970s, because that was the period in which group polarization was studied most extensively, and because the principal findings, for our purposes, have not been challenged in the subsequent literature (with several exceptions to be noted below).

⁶² Note that conformity does not explain group polarization. People are not attempting to conform, even under the social comparison theory; they are attempting to maintain their relative position, and the revelation of the views of others shifts people's conception of what judgment is necessary to maintain that position. See Myers, *supra* note, at 562, indicating that people "want to perceive themselves as somewhat different from others" and that "people want to differentiate themselves from others, to a small extent and in the right direction."

⁶³ There is an obvious connection between this point and recent work in economics on reputational influences on behavior. See Timur Kuran, *Private Truths, Public Lies* (1996). Note, however, that the social comparison approach emphasizes presentation of self to self, as well as self to others. "By 'one-upping' the self-presentations of others, people can see and present themselves as basically similar, yet desirably distinctive." Myers, *supra*, at 562.

⁶⁴ On signalling generally, see Eric Posner, *Symbols, Signals, and the Law* (forthcoming 2000).

desire.⁶⁵ No one can know what such a position would be until the positions of others are revealed.⁶⁶ Thus individuals move their judgments in order to preserve their image to others and their image to themselves. A key claim here is that information alone about the actual positions of others—without discussion—will produce a shift, and evidence has confirmed this fact. Research shows that this effect helps explain a shift toward caution (the “cautious shift”) as well.⁶⁷

The second explanation emphasizes the role of *persuasive arguments*.⁶⁸ The key point here is that an individual’s choice or position on an issue is a function of the number and persuasiveness of arguments presented. The choice therefore moves in the direction of the position held by the most confident and outspoken members of the group. Because a group that is inclined in a certain direction will have a disproportionate number of arguments supporting that direction, the result of discussion will be to move individuals further in the direction of their initial inclinations.

The persuasive arguments theory begins with the intuition that if a group is deliberating about some difficult question with a factual answer (how many countries are there in Africa, for example), discussion will produce some movement, not toward the mean, but toward the minority view on which one or a few members have

⁶⁵ For a quite vivid demonstration of such a process in the enactment of the Clean Air Act, one that does not, however, identify the mechanisms, discussed here, see Bruce Ackerman, John Millian, and Donald Elliott, *Toward a Theory of Statutory Evolution: The Federalization of Environmental Law*, 1 J. L. Econ. & Organization 313 (1985).

⁶⁶ “Once the real locations of the mean was known, should it not be the case, granting that everyone wanted to see himself as reasonably audacious, that those who were really below the mean would be motivated to adopt riskier positions and so change the mean and produce the risky shift?” Brown, *supra*, at 214.

⁶⁷ Investigations of social influence have emphasized both one-upmanship and the removal of pluralistic ignorance, that is, ignorance of what other people think (or are willing to say they think). Note that it is implicit in these findings that people seem to want not to conform, but to be different from others in a desirable way. “To be virtuous . . . is to be different from the mean – in the right direction and to the right degree.” Brown, *supra* note, at 469.

⁶⁸ There is an obvious connection between this theory and recent work in economics on informational influences on behavior and in particular on information cascades. See David Hirshleifer, *The Blind Leading The Blind*, in *The New Economics of Human Behavior* (1997).

accurate information. Of course many of the questions involving group polarization do not have purely factual answers. But a key aspect of those discussions is that the person with the correct answer is likely to state his view with a high degree of confidence, and also be able to make some argument in favor of that view. Novel arguments, bringing up fresh points, are especially likely to be persuasive. Thus it is suggested that the “important thing that happens in discussion is that individual arguments are expressed and become fully shared.”⁶⁹

When people hear arguments that they perceive as valid, or find to be new, they will shift in the direction suggested by those arguments. Discussion produces larger “argument pools.” If a group of moderately feminist women become more feminist, a group moderately opposed to gun control more extremely so, and so forth, the reason is that the “argument pool” of any such group will contain a preponderance of arguments in the direction suggested. Once the set of individual arguments is exposed to all individual members, there will be an inclination in the direction of initial inclinations, but more extremely and intensely so. “Because the choice dilemmas have total argument pools in which the balance favors either risk or caution, the larger sample of the pool made available to all in discussion will produce either a risky shift or a cautious shift according to the direction of prediscussion inclinations as revealed in the means of initial decisions.”⁷⁰ The suggestion is that group polarization will occur when convincing arguments produce a shift in the direction of prediscussion inclinations, revealed in the means of the initial decisions. At the same time, the persuasive arguments theory suggests that there will be “depolarization” if and when new persuasive arguments are offered that are opposite to the direction initially favored by group members, and there is evidence for this phenomenon as well.⁷¹

⁶⁹ Brown, *supra*, at 219.

⁷⁰ Brown, *supra*, at 219.

⁷¹ A third possibility is that hearing other similar opinions produces greater confidence in individual positions, opening members to a more extreme judgment in the same direction. raised recently by Heath and Gonzales . See Chip Heath and Richard Gonzales, Interaction With Others Increases Decision Confidence But Not Decision Quality: Evidence Against Information

Of course it is not the case that all groups polarize. There are two qualifications to the claims made thus far. First, these are statistical regularities, and some groups will not polarize at all, a statement confirmed by many individual juries in our study. Second, depolarization, rather than polarization, will be found when the relevant group consists of individuals drawn equally from two extremes.⁷² Thus if people who initially favor caution are put together with people who initially favor risk-taking, the group judgment will move toward the middle. It is remarkable, then, that despite the presence of some groups who did the opposite, the groups in our study who did polarize were numerous enough to cause the aggregate pattern of our results to mimic polarization.

V. Implications and Reforms

A. Is There A Problem? What Kind of Problem?

The group polarization phenomenon (and other evidence involving group performance) raises questions about the common belief that groups, and in particular juries, generally make better decisions than individuals. The common belief does seem sensible if only because the averaging of multiple judgments should reduce errors and variance—even if group deliberation has no effect on individual judgments.⁷³ But a now considerable literature, using a wide variety of tasks, concludes that deliberating groups hold no generalized advantage over individuals in the performance of judgment tasks.⁷⁴

More specifically, a review of judgmental biases in legal contexts suggests that jury deliberations are actually slightly more likely to amplify the biases of individuals than to attenuate them.⁷⁵ Indeed, a series of studies involving juries suggests that in a variety of contexts,

Collection Views Of Interactive Decision Making, 61 *Organizational Behavior and Human Decision Processes* 305-326 (1997).

⁷² See H. Burnstein, *Persuasion As Argument Processing*, in *Group Decision Making* (H. Brandstetter, J.H. Davis, and G. Stocker-Kreichgauer eds., 1982).

⁷³ See the discussion of Saks, *supra* note.

⁷⁴ Norbert Kerr, Robert MacCoun, and Geoffrey Kramer, *Bias in Judgment: Comparing Individuals and Groups*, 103 *Psychological Review* 687-719 (1996).

⁷⁵ *Id.*

jury judgments are likely to be more erratic than the judgments of (the mean or median of a group of the same number of) individuals.⁷⁶ The simplest conclusion from our study here is that to the extent that there is a concern about unpredictable damage awards, deliberation is not likely to alleviate that concern. On the contrary, deliberations increase unpredictability in the particular sense that the dollar judgment of one group of six (or twelve) is highly unlikely to predict the dollar judgment of another group of six (or twelve)—in fact it is less likely to predict that judgment than it would if the group judgment were based on the median judgment of the individual members.⁷⁷

Unpredictability is a serious problem for jury judgments, partly because it ensures that the similarly situated will not be treated similarly (and thus produces unfairness for plaintiffs and defendants alike), partly because it may produce overdeterrence in risk-averse defendants, partly because of the sheer cost involved in litigation-related expenses. But unpredictable awards might be worthwhile if many or most of them are sound; predictable awards are nothing to

⁷⁶In the jury damage assessment context, Diamond and Casper's (1992) mock jury awards were significantly higher than the mean individual award (by 26%). See Shari Diamond and Jonathan Casper, *Blindfolding the Jury to Verdict Consequences: Damages, Experts, and the Civil Jury*, 26 *Law and Society Review* 513-563 (1992). This finding is mirrored qualitatively in other studies, see James H. Davis, *Group Decision Making and Quantitative Judgments: A Consensus Model*, in *Understanding Group Behavior: Consensual Action by Small Groups* (E. Witte and J. Davis eds. 1996); Martin Kaplan and Charles Miller, *Group Decision Making and Normative Versus Informational Influence: Effects of Type of Issue and Assigned Decision Rule*, 53 *Journal of Personality and Social Psychology* 306 (1987) (for compensatory but not punitive damages). If this general trend applies to our previous study, see Sunstein, Kahneman, and Schakde, *supra*, at 2100-2104, then the synthetic, statistical "juries" analyzed there will have underestimated both the mean and the variance of the awards that deliberating juries would have made. Consistent with this conjecture, Gigone and Hastie, *supra*, found that group judgments have higher error variances than the average judgment of group members.

⁷⁷Note also that because our study stipulated compensatory damages, and held them fixed across cases, we may well have understated true variance in punitive awards, because according to previous research, see Eisenberg, *supra* note; Karpoff and Lott, *supra* note, real juries anchor on their own compensatory award, rather than into some constant value

celebrate if they are invariably too low or too high. Indeed, it is possible that the deterrent signal created by a range of variable awards, including a number of very high awards, is better, from the standpoint of preventing undesirable conduct, than the comparable signal of any realistic alternative.⁷⁸ There is thus an important remaining question: whether jury deliberation, including group polarization, produces a better deterrent signal or more accurate judgments than alternative arrangements.

It is certainly possible that the median judgment of a set of (say) 100 deliberating juries is more accurate than the median judgments of (say) 600 individuals. Without an independent account of good outcomes, in this contested area of the law, this possibility is very hard to assess.⁷⁹ But an understanding of the mechanisms that underlie group polarization certainly does not give much reason for optimism, at least in the context of damage awards. To the extent that polarization occurs as a result of social comparison, there is no reason to think that the higher and lower post-deliberative awards are better (or for that matter worse) than the lower and higher median of pre-deliberative individual judgments. When social comparison helps generate both larger and smaller awards, no relevant information is producing improved judgments.⁸⁰ The increased unpredictability caused by polarization, however, is still present and problematic, even if the means are no more or less valid.

To the extent that polarization occurs as a result of persuasive arguments, there might appear to be more reason to believe that group discussion produces better, more informed outcomes. Recall that with respect to purely factual questions on which one or a few

⁷⁸ Cf. Galanter and Luban, *supra* note.

⁷⁹ Compare the interesting claim that there is a "leniency shift" in criminal juries and that this shift produces more accurate judgments, because juries are more likely than individual jurors to apply the reasonable doubt standard correctly. See Robert MacCoun and Norbert Kerr, *Asymmetric Influence in Mock Jury Deliberation: Jurors' Bias for Leniency*, 54 *J Personality and Social Psych.* 21 (1988). There is no obvious analogy here, because it is not easy to find a legal standard against which to evaluate the increase in large awards and the decrease in small awards.

⁸⁰ To be sure, discussion does dissipate pluralistic ignorance, by showing people what others think; but this is hardly a guarantor of better outcomes, whatever may be our criteria for defining them.

group members have expertise, people are pushed toward a (correct) minority view simply because one or a few members of the group actually know the right answer.⁸¹ But it is not at all clear that there is an analogy in the context of punitive damage awards. In this setting, arguments that turn out to be persuasive may or may not be sound. This is especially so in light of the pervasive problem of “scaling without a modulus.” Even those with convincing arguments are likely not to have a good reason to favor one or another dollar figure (which is not to say that their numbers are worse than those of anyone else). We conclude that the fact that high awards become higher, and low awards become lower, should not be seen as a tribute to the power of deliberation to lead people in better directions—though aside from exacerbating the problem of unpredictability, it may not lead them in worse directions either.

B. Reform Proposals

With respect to punitive damage awards, and damages awards in related areas, many proposals have been motivated by a desire to decrease unpredictability. This goal has, for example, played a role in proposals for damage caps, for simple multipliers (relating punitive awards to compensatory awards), and for informing the jury of average awards or of intervals.⁸² It has played a role in constitutional limitations as well.⁸³ The desire to reduce unpredictability motivated our discussion of the possibility of eliciting from the jury not dollar awards, but normative judgments on a bounded numerical scale.⁸⁴ These judgments might be converted into a dollar award through some kind of calibration function, based on experts (“technocratic populism”) or on population-wide data relating normative judgments to dollar awards (“predictable populism”).⁸⁵ Of course it is possible to question whether it is practical to ask a deliberating jury to make a moral judgment on a bounded numerical scale, hardly an ordinary practice in daily life, and indeed a task that might seem even odder

⁸¹ Compare the finding of a “leniency shift” in MacCoun and Kerr, *supra* note.

⁸² See Saks et al., *supra* (discussing alternative approaches and their effects).

⁸³ See *BMW v. Gore*, 116 S. Ct. 1589, 1601 (1996) (Breyer, J., concurring).

⁸⁴ See Sunstein, Kahneman, and Schkade at 2112-2121.

⁸⁵ See *id.*

than the somewhat more familiar one of punishing wrongdoers through dollar awards.

The findings here do not lead directly to any particular reform proposal⁸⁶; but they add two points to the existing literature. First, they demonstrate that juries can use a bounded punishment scale reliably. Juries are able to answer the normative question directly, and they are also able to use a bounded numerical scale far more reliably than the familiar dollar scale. And if deliberating juries are thought to have advantages over other, less populist institutions—as many people clearly believe⁸⁷—then there is reason to consider a reform proposal that would involve directly eliciting the jury’s moral judgment. As noted, these judgments might be converted into a dollar award by some kind of calibration formula, defined by expert judgments about what different dollar awards would mean or do to particular defendants, or instead on population-wide data relating normative judgments to dollar awards. Either route would greatly diminish unpredictability.⁸⁸ The data here, along with previous data, show that a calibration formula is also feasible to develop and use.⁸⁹

The second point stems from the finding that deliberating juries do not reduce erratic awards, and indeed that the process of deliberation is one of the causes of unpredictability. We have seen that deliberation can even produce “runaway juries,” if these are understood as juries in which group discussion yields awards much higher than those of even the highest of pre-discussion judgments. Such juries can be produced even in a mock jury task. Thus our findings fortify the suggestion that difficulties with the dollar scale make it hazardous to continue to rely on the current system, in which juries must map their moral judgments onto that scale without being given any guidance about the meaning of the various “points” on the scale.

⁸⁶ They do suggest, somewhat amusingly, a simple way to promote predictability: Do not allow jurors to deliberate, and simply take the mean of their individual damage judgments! We do not discuss this alternative, however, because it is so foreign to our traditions, and also because a great deal of unpredictability would remain, as suggested by our data here, see Table 4, and see also SKS at 2100-2104.

⁸⁷ See Galanter and Luban, *supra* note.

⁸⁸ See Sunstein, Kahneman, and Schkade, at 2113-2120.

⁸⁹ *Id.* at 2112-2118.

To be sure, eliciting moral judgments rather than dollar awards would not answer all of the relevant questions, because group polarization also produces not only higher and lower dollar awards, but also higher and lower moral judgments. There is no reason, however, to think that the resulting “verdicts” are worse than the mean of individual judgments before discussion, just as there is no reason to think that they are better (except that increased unpredictability is unambiguously worse in itself). In the absence of an independent criterion of rightness, it probably makes sense to continue the tradition of obtaining the deliberative judgment, without at the same time creating the forms of unpredictability and unfairness that can come from the unbounded dollar scale.

The most radical reform would be to dispense with the jury entirely and to move toward judicial judgments or even to develop a kind of penalty schedule, based on the judgments of some combination of representative and expert institutions.⁹⁰ We cannot evaluate these alternatives here. But to the extent that our study shows both decreasing and increasing awards as a result of group polarization, operating independently of better-informed or sounder judgments of any kind, it supplies additional support for that more radical reform. Of course the radical reform might be rejected if the relevant institutions would be unreliable, perhaps because bureaucracies might be vulnerable to the exercise of politically powerful private groups. The question is one of comparative institutional competence. What we have added here is that the process of deliberation will increase high awards and diminish low awards, a result that cannot be comforting in light of the increased unpredictability and the pervasive problem of scaling without a modulus.

⁹⁰ The idea has received considerable attention in the analogous area of contingent valuation. See Murray Rutherford et al., *Assessing Environmental Losses: Judgments of Importance and Damage Schedules*, 22 *Harv. Envtl. L. Rev.* 51 (1998); Richard B. Stewart, *Liability for Natural Resource Injury*, in *Analyzing Superfund 219*, 241-44 (Richard L. Revesz and Richard B. Stewart eds. 1995). In the area of compensatory damages, see the plea for damages schedules in Bovbjerg et al., *supra*. In the punitive damage context, see Viscusi, *supra* note; Sunstein, Kahneman, and Schkade, *supra*, at 2121-2126. For damages generally, see Atiyah, *supra* note.

VI. Conclusion

We have found that as compared with the median of individual judgments, deliberation substantially increases group polarization: Small awards become smaller, and large awards become (significantly) larger. The point has implications for damage awards in general and also for understanding social deliberation. From the normative point of view, it is hard to know whether the resulting judgments are better than the median of pre-deliberative individual judgments. But four points seem clear. First, moral judgments about personal injury cases are very widely shared over diverse communities and demographic categories. Second, those shared moral judgments do not produce predictable dollar awards. Third, group polarization occurs, in a quite dramatic fashion, in the context of damage awards. Hence both social influence and persuasive arguments drive group judgments to more extreme points in the direction of the inclination originally indicated by the median of pre-deliberation judgments. Fourth, the problem of unpredictable and erratic judgments is likely to be aggravated, rather than alleviated, by virtue of the fact that juries are deliberative bodies.

Appendix

Glover v. General Assistance

Joan Glover, a five-year-old child, ingested a large number of a non-prescription allergy medicine called Allerfree, and required a three-week hospital stay. The Allerfree bottle used a faulty childproof safety cap. The Glovers sued the manufacturer of Allerfree, the General Assistance company. The trial jury ordered General Assistance to pay the Glovers \$200,000 in compensatory damages.

Facts of the Case Established at Trial

Joan's parents testified that after her birth they had "childproofed" their house and ensured that all of their medications had childproof safety caps. The Allerfree bottle carries a label reading "Childproof Cap." Joan found the pills in a kitchen drawer and ingested most of the bottle. The overdose permanently 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.

General Assistance is a large company (with profits of \$100-200 million per year) that manufactures a variety of non-prescription medicines. The company has sold tens of thousands of bottles of medicines with childproof safety caps that were generally effective, but had a failure rate much higher than any others in the industry. Internal company documents showed that General Assistance had chosen to ignore federal regulations requiring more effective safety caps. An internal memo presented at trial says that "this stupid, unnecessary federal regulation is a waste of our money"; it acknowledges the risk that Allerfree might be punished for violating the regulation but says "the punishments 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 General Assistance executive that the company was "on shaky ground on this one."

Closing Argument by Glovers' Attorney

The attorney for the Glovers argued that General Assistance's disregard for children's safety and for the law was abhorrent and represented exactly the kind of reckless corporate greed deserving of a high award of punitive damages. He concluded that General Assistance's shocking profit-mongering should be punished so that the company would not feel itself at liberty to put children at risk in the future.

Closing Argument by General Assistance's Attorney

The attorney for General Assistance emphasized that while the cap had a high failure rate relative to others on the market, it had nonetheless been conceded at trial that the cap was effective in most cases. She argued that, given that the FDA official had only communicated to them verbally, and had not required the company to take any action, it was not at all clear that the cap was actually in violation of the regulation at all.

This is a preliminary draft of a paper forthcoming in the *Columbia Law Review*. All rights reserved.

Readers with comments should address them to:

Cass R. Sunstein
Karl N. Llewellyn Distinguished Service Professor
University of Chicago Law School
1111 East 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).
51. William M. Landes, Lawrence Lessig, and Michael E. Solimine, Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges (January 1998).

52. John R. Lott, Jr., A Simple Explanation for Why Campaign Expenditures are Increasing: The Government is Getting Bigger (February 1998).
53. Richard A. Posner, Values and Consequences: An Introduction to Economic Analysis of Law (March 1998).
54. Denise DiPasquale and Edward L. Glaeser, Incentives and Social Capital: Are Homeowners Better Citizens? (April 1998).
55. Christine Jolls, Cass R. Sunstein, and Richard Thaler, A Behavioral Approach to Law and Economics (May 1998).
56. John R. Lott, Jr., Does a Helping Hand Put Others At Risk?: Affirmative Action, Police Departments, and Crime (May 1998).
57. Cass R. Sunstein and Edna Ullmann-Margalit, Second-Order Decisions (June 1998).
58. Jonathan M. Karpoff and John R. Lott, Jr., Punitive Damages: Their Determinants, Effects on Firm Value, and the Impact of Supreme Court and Congressional Attempts to Limit Awards (July 1998).
59. Kenneth W. Dam, Self-Help in the Digital Jungle (August 1998).
60. John R. Lott, Jr., How Dramatically Did Women's Suffrage Change the Size and Scope of Government? (September 1998)
61. Kevin A. Kordana and Eric A. Posner, A Positive Theory of Chapter 11 (October 1998)
62. David A. Weisbach, Line Drawing, Doctrine, and Efficiency in the Tax Law (November 1998)
63. Jack L. Goldsmith and Eric A. Posner, A Theory of Customary International Law (November 1998)
64. John R. Lott, Jr., Public Schooling, Indoctrination, and Totalitarianism (December 1998)
65. Cass R. Sunstein, Private Broadcasters and the Public Interest: Notes Toward A "Third Way" (January 1999)
66. Richard A. Posner, An Economic Approach to the Law of Evidence (February 1999)
67. Yannis Bakos, Erik Brynjolfsson, Douglas Lichtman, Shared Information Goods (February 1999)

68. Kenneth W. Dam, Intellectual Property and the Academic Enterprise (February 1999)
69. Gertrud M. Fremling and Richard A. Posner, Status Signaling and the Law, with Particular Application to Sexual Harassment (March 1999)
70. Cass R. Sunstein, Must Formalism Be Defended Empirically? (March 1999)
71. Jonathan M. Karpoff, John R. Lott, Jr., and Graeme Rankine, Environmental Violations, Legal Penalties, and Reputation Costs (March 1999)
72. Matthew D. Adler and Eric A. Posner, Rethinking Cost-Benefit Analysis (April 1999)
73. John R. Lott, Jr. and William M. Landes, Multiple Victim Public Shooting, Bombings, and Right-to-Carry Concealed Handgun Laws: Contrasting Private and Public Law Enforcement (April 1999)
74. Lisa Bernstein, The Questionable Empirical Basis of Article 2's Incorporation Strategy: A Preliminary Study (May 1999)
75. Richard A. Epstein, Deconstructing Privacy: and Putting It Back Together Again (May 1999)
76. William M. Landes, Winning the Art Lottery: The Economic Returns to the Ganz Collection (May 1999)
77. Cass R. Sunstein, David Schkade, and Daniel Kahneman, Do People Want Optimal Deterrence? (June 1999)
78. Tomas J. Philipson and Richard A. Posner, The Long-Run Growth in Obesity as a Function of Technological Change (June 1999)
79. David A. Weisbach, Ironing Out the Flat Tax (August 1999)
80. Eric A. Posner, A Theory of Contract Law under Conditions of Radical Judicial Error (August 1999)
81. David Schkade, Cass R. Sunstein, and Daniel Kahneman, Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages (September 1999)