6-2002

Reconciling Experimental Incoherence with Real-World Coherence in Punitive Damages

Theodore Eisenberg

Cornell Law School, ted-eisenberg@lawschool.cornell.edu

Jeffrey J. Rachlinski

Cornell Law School, rachlins@law.mail.cornell.edu

Martin T. Wells

Cornell University, mtw1@cornell.edu

Follow this and additional works at: <http://scholarship.law.cornell.edu/facpub>

Part of the [Applied Statistics Commons](#), [Civil Procedure Commons](#), [Law and Society Commons](#), and the [Remedies Commons](#)

Recommended Citation

Eisenberg, Theodore; Rachlinski, Jeffrey J.; and Wells, Martin T., "Reconciling Experimental Incoherence with Real-World Coherence in Punitive Damages" (2002). *Cornell Law Faculty Publications*. Paper 377.
<http://scholarship.law.cornell.edu/facpub/377>

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Faculty Publications by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

Reconciling Experimental Incoherence with Real-World Coherence in Punitive Damages

Theodore Eisenberg,* Jeffrey J. Rachlinski**
& Martin T. Wells***

Experimental evidence generated in controlled laboratory studies suggests that the legal system in general, and punitive damages awards in particular, should display an incoherent pattern. According to the prediction, inexperienced decisionmakers, such as juries, should fail to convert their qualitative judgments of defendants' conduct into consistent, meaningful dollar amounts. This Article tests this prediction and finds modest support for the thesis that experience across different types of cases will lead to greater consistency in awards. Despite this support, numerous studies of damage awards in real cases detect a generally sensible pattern of damage awards. This Article tries to reconcile the largely coherent pattern of real-world results with the experimental findings and suggests that careful attention to sources of coherence and incoherence can help reconcile experimental and real-world results.

Few complaints about a legal system resonate louder than charges of incoherence.¹ A system that fails to treat similarly situated parties equally cannot be squared with fundamental notions of fairness and justice.² Incoherence unintentionally converts law into a lottery. Even charges of bias are less pernicious in some ways, because the targets of bias can organize and voice their opposition. Incoherence has no constituency and yet also no organized opposition.

* Henry Allen Mark Professor of Law, Cornell Law School. We would like to thank participants at the American Law and Economics Association annual meeting on May 5-6, 2002, in Cambridge, Massachusetts, for their comments.

** Professor of Law, Cornell Law School.

*** Professor of Statistics, Department of Social Statistics, and Elected Member of the Law Faculty, Cornell University.

1. Charges of incoherence are one of the two principal claims of the critical legal studies school (bias being the other). See MARK G. KELMAN, A GUIDE TO CRITICAL LEGAL STUDIES (1987); Mark G. Kelman, *Trashing*, 36 STAN. L. REV. 293 (1984).

2. See Cass R. Sunstein, Daniel Kahneman & David Schkade, *Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)*, 107 YALE L.J. 2071, 2077 (1998) ("If similarly situated people—plaintiffs and defendants alike—are not treated similarly, erratic awards are unfair.").

Recent psychological research findings of incoherence pose a dilemma for law. Cognitive psychologists suggest that human judgment is, by nature, incoherent.³ Judgments made in an individual case often do not comport with judgments made in the aggregate. In particular, *Predictably Incoherent Judgments* ("PIJ") provides convincing experimental evidence that people lack the basic cognitive skills necessary to translate qualitative moral judgments into quantitative numeric scales.⁴ To the extent that a legal system relies on individuals to accomplish such translations, it risks incorporating incoherence into its judgments.

Concern that the legal system requires people to make judgments that are apt to be unreliable is common enough, but the ubiquitousness of the type of judgments that PIJ identifies raises the stakes by orders of magnitude. Converting moral judgments into quantitative scales, such as dollars or years in prison, is a ubiquitous task for law. No legal system, whether it operates through the courts, administrative agencies, or acts of Congress, can function without being able to make such conversions effectively. With a few exceptions,⁵ the law is expressed through dollars or years of imprisonment. If damage awards and prison sentences are "predictably incoherent," then our legal system (indeed, every legal system) might be so pervasively unfair as to be indefensible.

The psychological theories that PIJ relies on are consistent with decades of research in cognitive psychology in general, and are supported in particular by dozens of studies on the issue of how people translate qualitative scales into quantitative ones.⁶ Furthermore, its experimental translation of this research into the legal domain is also solid. The conclusion that people incoherently translate the moral into the numeric within the context of legal decisions is therefore hard to reject.

Therein lies the paradox. The legal system produces hundreds of quantitative judgments each day. If we take PIJ's results seriously, and accept its characterization of the situation as scandalous,⁷ should we not be observing a legal system on the verge of collapse? Any civil society would find a torrent

3. See Daniel Kahneman, *New Challenges to the Rationality Assumption*, 3 LEGAL THEORY 105, 106 (1997) ("Numerous experiments illustrate beliefs and preferences that violate a fundamental requirement [of rationality].").

4. Cass R. Sunstein, Daniel Kahneman, David Schkade & Ilana Ritov, *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153 (2002) [hereinafter PIJ]. Some of these authors make similar points in an earlier, related article. See Sunstein et al., *supra* note 2.

5. The law sometimes requires qualitative judgments as well, although these may suffer from similar forms of incoherence. See, e.g., Mark Kelman, Yuval Rootenstreich & Amos Tversky, *Context-Dependence in Legal Decision Making*, 25 J. LEGAL STUD. 287 (1996).

6. Sunstein et al., *supra* note 2, at 2107-08 (describing research on scaling without a modulus).

7. PIJ, *supra* note 4, at 1156, 1181.

of incoherent legal judgments to be unfair, unjust, unpredictable, and unmanageable. People would not be able to plan their lives, business would not be able to function, and government itself might be at risk. In the United States, however, a country that produces as many legal judgments per capita as most,⁸ the legal system functions reasonably, if not perfectly. Most people generally understand the legal consequences of such behaviors as driving too fast, buying a house, or purchasing illegal narcotics. Intuitively, the system seems to work well, despite the well-founded concerns in *PIJ*.

More important than intuition, however, is that the research on quantitative judgments demonstrates substantial consistency in judgments. The results of lawsuits appear to be quite rational. In civil lawsuits, the worst-behaved defendants who caused the most harm are the most likely to lose and pay the most damages.⁹ A recent review of the literature concludes that “numerous field studies support the notion that jury damage awards are at least moderately related to the seriousness of the plaintiff’s injury,” and that other “field studies have found that punitive damage award amounts are strongly related to the size of the compensatory damages awarded.”¹⁰ Most notably, given *PIJ*’s focus,

8. But the United States is not so litigious as most people believe. PATRICIA M. DANZON, *MEDICAL MALPRACTICE: THEORY, EVIDENCE, AND PUBLIC POLICY* 23-24 (1985) (stating that “at most 1 in 10 negligent injuries results in a claim”); DEBORAH R. HENSLEY, M. SUSAN MARQUIS, ALLAN F. ABRAHAMSE, SANDRA H. BERRY, PATRICIA A. EBENER, ELIZABETH G. LEWIS, E. ALLEN LIND, ROBERT J. MACCOUN, WILLARD G. MANNING, JEANNETTE A. ROGOWSKI & MARY E. VAIANA, *COMPENSATION FOR ACCIDENTAL INJURIES IN THE UNITED STATES* 121 (1991) (showing low rate of claims for various accident types); HARVARD MEDICAL PRACTICE STUDY, *PATIENTS, DOCTORS, AND LAWYERS: MEDICAL INJURY, MALPRACTICE LITIGATION, AND PATIENT COMPENSATION IN NEW YORK* 7-1 (1990) (“[W]e estimate that eight times as many patients suffer an injury from medical negligence as there are malpractice claims.”); Herbert M. Kritzer, *Lawyer Fees and Lawyer Behavior in Litigation: What Does the Empirical Literature Really Say?*, 81 TEXAS L. REV. (forthcoming 2002) (manuscript at 39, on file with author).

9. PATRICIA M. DANZON, *MEDICAL MALPRACTICE: THEORY, EVIDENCE, AND PUBLIC POLICY* 40-42 (1985); Henry S. Farber & Michelle J. White, *Medical Malpractice: An Empirical Examination of the Litigation Process*, 22 RAND J. ECON. 199 (1991); Evan Osborne, *Courts as Casinos? An Empirical Investigation of Randomness and Efficiency in Civil Litigation*, 28 J. LEGAL STUD. 187, 197-98 (1999); Frank A. Sloan & Chee Ruey Hsieh, *Variability in Medical Malpractice Payments: Is the Compensation Fair?*, 24 L. & SOC’Y REV. 997, 1019 (1990) (“In general, the results are consistent with the notion that claimants who have incurred higher losses receive more compensation.”); Neil Vidmar & Jeffrey J. Rice, *Assessments of Noneconomic Damage Awards in Medical Negligence: A Comparison of Jurors with Legal Professionals*, 78 IOWA L. REV. 883 (1993); Neil Vidmar, *The Performance of the American Civil Jury: An Empirical Perspective*, 40 ARIZ. L. REV. 849, 868-71 (1998).

10. Dennis J. Devine, Laura D. Clayton, Benjamin B. Dunford, Rasmy Seying & Jennifer Pryce, *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 PSYCHOL., PUB. POL’Y & L. 622, 706 (2001).

awards of punitive damages, an area that should be riddled with incoherence, produce remarkably consistent judgments.¹¹

Although psychological research as applied to law is sometimes charged with being unrealistic and inapplicable to the real world,¹² we do not suggest that the answer to this paradox lies largely in methodological difficulties with the application of cognitive psychology beyond the laboratory. We believe that *PIJ*'s experimental results should be taken seriously. Rather, the answer lies in the legal system's ability to reduce and manage the limitations of human judgment. If human judgment produces incoherence, it is likely that legal systems (like other professional systems such as medicine and engineering)¹³ identify ways to reduce the incoherence of such judgments or the effects of such incoherence. In other contexts, the law has managed to adapt to the cognitive biases of decisionmakers.¹⁴ Nevertheless, such adaptations might be improved upon in some ways. The question for reform-minded psychologists and the consumers of psychology, then, is when and where costly incoherence persists in a form in which it can be effectively remedied.

Identifying such circumstances requires more than theorizing and more than conducting further laboratory-oriented experiments; it requires a serious look at the real world of cases. Determining when and in what ways the cognitive sources of incoherent individual judgment produce incoherent legal judgments requires developing the implications of the psychological research for observed cases, and testing whether these implications hold true.

This Essay begins that testing process. Part I reviews the available data on punitive damages because, according to *PIJ*, such awards should be a prime area of incoherence. Despite this prediction, the available data from many sources, rather than demonstrating incoherence, reveal a fairly sensible pattern of results. Part II presents new research, using existing data, that nevertheless supports the influence of the cognitive sources of incoherence that *PIJ* identifies. So a largely coherent real-world system coexists with some empirical support for *PIJ*'s experimentally detected incoherence. This suggests that real-world incoherence is not the scandalous plague *PIJ* fears but that the system should still strive to reduce its more modest degree of actual incoherence. Part III tries to sort through the sources of coherence and

11. Theodore Eisenberg, John Goerdt, Brian Ostrom, David Rottman & Martin T. Wells, *The Predictability of Punitive Damages*, 26 J. LEGAL STUD. 623, 633-37 (1997).

12. Jennifer Arlen, *Comment: The Future of Behavioral Economic Analysis of Law*, 51 VAND. L. REV. 1765 (1998); Robert A. Hillman, *The Limits of Behavioral Decision Theory in Legal Analysis: The Case of Liquidated Damages*, 85 CORNELL L. REV. 717 (2000).

13. See Chip Heath et al., *Cognitive Repairs: How Organizational Practices Can Compensate for Individual Shortcomings*, 20 RES. IN ORG'L BEHAV. 1 (1998).

14. Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571 (1998) [hereinafter Rachlinski, *Judging in Hindsight*]; Jeffrey J. Rachlinski, *Heuristics and Biases in the Courts: Ignorance or Adaptation?*, 79 OR. L. REV. 61 (2000).

incoherence to reconcile *PIJ*'s experimental findings with its modest degree of support in real-world data. Part IV explores the implications of *PIJ*'s findings about incoherence.

I. INCOHERENCE NOT OBSERVED: THE PUZZLING PREDICTABILITY OF PUNITIVE AWARDS AND OTHER LEGAL DECISIONMAKING

If incoherence distorts the determination of punitive damages awards, then the real world ought to contain substantial evidence of the need for reform. The primary evidence should be either a crazy pattern of awards or many anecdotal awards that are plainly unjust. *PIJ* seems to simply assume that such evidence exists. It refers to the "notorious unpredictability of punitive damages awards," a proposition for which it cites only the authors' own experimental evidence.¹⁵ Yet, despite many well-funded articles favoring punitive damages reform,¹⁶ researchers have not identified either a crazy pattern of awards or a substantial series of actual punitive damage awards that constitute a shocking pattern of incoherence or unfairness.

The limited focus of punitive damages reform advocates on experiments and theory is telling, because those who do look at actual cases are impressed by the scarcity of punitive awards and their surprisingly coherent pattern. Juries rarely award punitive damages¹⁷ and generally confine punitive damage awards to cases involving intentional misconduct; punitive damage awards in products liability and medical malpractice cases are exceptionally rare.¹⁸

15. *PIJ*, *supra* note 4, at 1168 & n.55.

16. Reid Hastie & W. Kip Viscusi, *What Juries Can't Do Well: The Jury's Performance as a Risk Manager*, 40 ARIZ. L. REV. 901 (1998); Sunstein et al., *supra* note 2; W. Kip Viscusi, *Juries, Judges, and The Mistreatment of Risk by the Courts*, 30 J. LEGAL STUD. 107 (2001); W. Kip Viscusi, *The Challenge of Punitive Damages Mathematics*, 30 J. LEGAL STUD. 313 (2001). See also Elizabeth Amon, *Exxon Bankrolls Critics of Punitives Then it Cites The Research in Appeal of \$5.3b Valdez Award*, NAT'L L.J., May 17, 1999, at A1. Some of this work is criticized in Richard Lempert, *Juries, Hindsight, and Punitive Damage Awards: Failures of a Social Science Case for Change*, 48 DEPAUL L. REV. 867 (1999).

17. See STEPHEN DANIELS & JOANNE MARTIN, *CIVIL JURIES AND THE POLITICS OF REFORM* (1995); Thomas A. Eaton, Susette M. Talarico & Richard E. Dunn, *Another Brick in the Wall: An Empirical Look at Georgia Tort Litigation in the 1990s*, 34 GA. L. REV. 1049, 1094 (2000) ("Punitive damages currently are not a significant factor in personal injury litigation in Georgia."); Theodore Eisenberg, Neil LaFountain, Brian Ostrom, David Rottman & Martin T. Wells, *Juries, Judges, and Punitive Damages: An Empirical Study*, 87 CORNELL L. REV. 743 (2002); Eisenberg et al., *supra* note 11, at 633-37 (summarizing studies); Neil Vidmar & Mary R. Rose, *Punitive Damages by Juries in Florida: In Terrorem and In Reality*, 38 HARV. J. LEGIS. 487, 487 (2001) (concluding that, "in Florida . . . the frequency of punitive damages was strikingly low").

18. See Eisenberg et al., *supra* note 11, at 635-37 (summarizing studies); Deborah Jones Merritt & Kathryn Ann Barry, *Is the Tort System in Crisis? New Empirical Evidence*, 60 OHIO ST. L.J. 315, 388 (1999) (finding no punitive awards in medical malpractice or

Judges behave similarly.¹⁹ When either juries or judges do award punitive damages, they do so in similar ways that relate strongly to compensatory awards.²⁰

Those who study actual cases searching for large numbers of unjust anecdotal awards come away more impressed by the justifiability of the awards.²¹ Even alleged hotbeds of punitive damages abuse, when subjected to social science scrutiny, turn out to be more sober than shocking. For example, Alabama juries, occasional poster-children for punitive-damages reform, award punitive damages in financial injury cases at rates similar to those of other states, and their punitive damage awards are generally smaller than those in other states.²²

Yet the deeply ingrained belief in the craziness of punitive awards is so profound that each time a study reveals the overall sober pattern of awards the news media treat it as a major new discovery.²³ In many respects, the public's apparent failure to appreciate the coherence of punitive damage awards

products liability cases in a twelve-year period in Franklin County, Ohio); Vidmar & Rose, *supra* note 17 (stating that in Florida study, "with the exception of asbestos cases, punitive damages were almost never given in products liability cases"). J. Clark Kelso & Kari C. Kelso, *An Analysis of Punitive Damages in California Courts, 1991-2000*, at 8 (unpublished manuscript), reports 11 products liability punitive damages awards in ten years of California litigation; however, they acknowledge that their data sample "underestimates" the number of punitive awards. *Id.* at 2.

19. Eisenberg et al., *supra* note 17.

20. Eisenberg et al., *supra* note 11; at 637-39, 647-52; Eisenberg et al., *supra* note 17; Theodore Eisenberg & Martin T. Wells, *The Predictability of Punitive Damages Awards in Published Opinions, the Impact of BMW v. Gore on Punitive Damages Awards, and Forecasting Which Punitive Awards Will Be Reduced*, 7 SUP. CT. ECON. REV. 59 (1999) [hereinafter Eisenberg & Wells, *Predictability on Appeal*]; Theodore Eisenberg & Martin T. Wells, *Punitive Awards After BMW, a New Capping System, and the Reported Opinion Bias*, 1998 WIS. L. REV. 387 [hereinafter Eisenberg & Wells, *Reported Opinions*]; Erik K. Moller, Nicholas M. Pace & Stephen J. Carroll, *Punitive Damages in Financial Injury Jury Verdicts*, 28 J. LEGAL STUD. 283, 300 n.52 (1999); Jonathan M. Karpoff & John R. Lott, Jr., *On the Determinants and Importance of Punitive Damage Awards*, 42. J. LAW & ECON. 527, 543 (1999); Vidmar & Rose, *supra* note 17, at 501 tbl. 3 (showing that thirteen years of Florida jury punitive awards had a median punitive-compensatory ratio of 0.7 to 1). For experimental evidence of similar punitive damages behavior by judges and jury-eligible citizens, see Jennifer K. Robbennolt, *Punitive Damages Decision Making: The Decisions of Citizens and Trial Court Judges*, 26 J.L. & HUM. BEHAV. 315, 327, 328 (2002).

21. See, e.g., WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 185 (1987); Michael Rustad & Thomas Koenig, *Reconceptualizing Punitive Damages in Medical Malpractice: Targeting Amoral Corporations, Not "Moral Monsters,"* 47 RUTGERS L. REV. 975, 1027, 1030 n.168 (1995).

22. Moller et al., *supra* note 20, at 333-34.

23. See, e.g., William Glaberson, *A Study's Verdict: Jury Awards Are Not Out of Control*, N.Y. TIMES, Aug. 6, 2001, at A9; Edward Felsenthal, *Punitive-Damage Awards Found To Be Generally Modest and Rare*, WALL ST. J., June 17, 1996, at B4.

resembles a classic example of the availability heuristic.²⁴ The availability heuristic refers to the tendency for people to base their understanding of the likelihood of uncertain events on the ease with which they can call examples of the events to mind.²⁵ As a consequence of this heuristic, people overestimate the occurrence of events that are reported in the media or that are particularly vivid and memorable.²⁶ Furthermore, “availability entrepreneurs” try to draw attention to such events so as to move public opinion in particular directions.²⁷ In the case of punitive damages, the media, perhaps with a nudge by availability entrepreneurs, reports “outrageous” damage awards, such as the infamous but largely explicable \$2.9 million damages award against McDonald’s for coffee that was too hot.²⁸ Left unreported are the mass of sensible awards (not to mention the inevitable appeals and settlements that bring high awards into line with those of other cases). This bias might even affect the managers of large companies, who constantly agitate for damages reform. Even if the public can be excused for reliance on a simple heuristic instead of hard data, the sober pattern of punitive awards should no longer be news to scholars with access to the data.

In short, virtually all of the serious empirical evidence portrays a sober, largely coherent system of punitive damages awards, a fact that has been well known for years. One challenge is to reconcile *PIJ*’s theoretical assertions and experimental findings with that mass of data.

II. INCOHERENCE OBSERVED: TESTING *PIJ*’S THESIS USING ACTUAL CASES

PIJ’s concept of category-boundedness as a source of incoherence interests both experimental and real-world analysts. This shared interest offers an

24. See generally Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1127-38 (1974).

25. See *id.*

26. Paul Slovic, Baruch Fischhoff & Sarah Lichtenstein, *Facts Versus Fears: Understanding Perceived Risk*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982) (demonstrating a correlation between the rate at which disasters are reported and the degree of overestimation of the likelihood that such events occur).

27. Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683 (1999) (discussing the effect of the availability heuristic on public opinion).

28. *Liebeck v. McDonald’s Restaurants, P.T.S., Inc.*, No. CV 93-02419, 1995 WL 360309 (D.N.M. Aug. 18, 1994). The jury awarded \$160,000 in compensatory damages for severe burns suffered by the plaintiff and \$2,700,000 in punitive damages. The trial judge reduced the punitive award to \$480,000. The case was ultimately settled, presumably for a lesser amount still. See Neil Vidmar, Felicia Gross & Mary Rose, *Jury Awards for Medical Malpractice and Post-Verdict Adjustments of Those Awards*, 48 DEPAUL L. REV. 265, 266 (1998); Andrea Gerlin, *A Matter of Degree: How a Jury Decided that a Coffee Spill Is Worth \$2.9 Million*, WALL ST. J., Sept. 1, 1994, at A1.

opportunity to use actual case data to help illuminate theoretical and experimental insights into incoherence across categories.

A. *Generating Testable Hypotheses Using the PIJ Results*

PIJ uses categories as a loosely defined instance of classification. “Category-bound” thinking can lead to differences in treatment across categories, a source of injustice.²⁹ For present purposes, it is important that categories, as used in *PIJ*, include case categories commonly used to designate the subject matters of lawsuits—categories such as torts, fraud, and the like. This is shown by *PIJ*’s use of case categories to illustrate the problem of decisionmaking across categories.³⁰ *PIJ* describes “personal injury cases, business fraud cases, sexual harassment cases or libel cases” as categories within each of which, “diverse people are likely to agree on how to rank” punishment-worthiness.³¹ When one moves from within to across case categories, however, *PIJ* hypothesizes that category-bound thinking leads to large shifts in punitive intentions and actual punishments, to what *PIJ* labels judgment reversals, leading to an unacceptable breakdown of coherence.³²

PIJ regards punitive damages decisionmaking as a troublesome instance of category-bound thinking leading to incoherence: “[O]utrage and its manifestations—including punitive damages—involve a process that situates any particular case in relation to its category. We will show that this manifestation of category-bound thinking yields systematic incoherence in punitive intentions and in punitive awards.”³³

And *PIJ* presents experimental evidence that category-bound thinking leads to troublesome variation in punitive awards.

The generality of *PIJ*’s claim is noteworthy. Unlike other critiques of punitive damages, *PIJ*’s concern is not an unsupported rant that punitive damages are out of control or a critique of punitive damages based on outlier behavior. Much concern about punitive awards focuses on actual or perceived outliers—the occasional crazy jury, or the poor, rural county in Alabama, Texas, or Mississippi that is said to be a magnet for punitive awards, that

29. *PIJ*, *supra* note 4, at 1156.

30. *Id.* at 1158 (using “personal injury cases, business fraud cases, sexual harassment cases or libel cases” to illustrate different categories, within each of which greater coherence is expected than is expected across categories).

31. *Id.* We have some reservations about an ordinal ranking with personal-injury cases always ranked higher than financial injury cases. When insurance companies wrongfully refuse to pay for covered medical procedures, we expect outrage to be especially high. In fact, the ratio of punitive award size to compensatory award size in insurance company misbehavior cases is unusually high. Eisenberg & Wells, *Predictability on Appeal*, *supra* note 20, at 84 tbl.1.

32. *PIJ*, *supra* note 4, at 1158.

33. *Id.* at 1165.

frightens businesses into unfair settlements. *PIJ*'s claim of incoherence is not limited to outliers; it bears directly on the mass of punitive damages awards.

Our challenge is to relate *PIJ*'s experimental category-based incoherence claims to what one observes in the real world of punitive damages. One possibility is to explore differences in punitive damages behavior across case categories. *PIJ* forecasts incoherence across categories, but how should that be observed in actual cases? This Part first identifies testable hypotheses generated by *PIJ*'s theorizing. It then describes the data used to test the hypotheses and reports the results, which modestly support *PIJ*'s theory and experimental findings.

We start with *PIJ*'s formula, which is said to capture "the psychological conception of punishment that we apply and extend in this article":³⁴

$$(1) \quad \text{Punishment (\$)} = \begin{array}{ccc} \text{Outrageousness} & \times & \text{Severity} & \times & \text{Translation} \\ & \text{of behavior} & & \text{of harm} & \text{factor} \end{array}$$

PIJ reports that people generally agree on Equation (1)'s first two terms, the outrageousness of behavior and the severity of harm.³⁵ The third term, the translation factor, generates the variability that leads to incoherence. Specifically, the translation factor ought to vary substantially or unpredictably across case categories. Within case categories, *PIJ* reports no incoherence problem.

Available data allow us to study whether the translation factor does indeed vary across case categories. One possibility is to observe variation in punitive awards across case categories, while controlling for Equation (1)'s non-translation-factor components. We use the best available measure of the severity of harm, the compensatory damages award in a case. For now, we ignore the influence of the outrageousness of behavior because of data limitations, and focus on the translation factor. This leads to the following model of the size of punitive damages awards:³⁶

$$(2) \quad \text{Punishment (\$)} = \text{Compensatory award (\$)} \times \begin{array}{c} \text{Translation} \\ \text{factor} \end{array}$$

To generate predictions from *PIJ* about this model, we begin with the observation that juries do not see multiple cases and hence have no opportunity to assess punitive award amounts in a context that allows comparison of awards

34. *Id.* at 1166.

35. *Id.*

36. Available data allow exploring punitive award patterns across case categories more reliably than exploring patterns within categories. This is because the most important case-level information is the case category. Available data do not allow study of within-category variation.

across categories. Juries address the question of punitive damages “in isolation”—one category at a time—rather than “in the context of a problem from a different category.”³⁷ Their situation is thus analogous to that of the first condition in *PIJ*’s experiments. Juries do not make awards with reference to other case categories. Their translation factor should therefore differ from that of judges.

B. Hypotheses

1. Case-category effects.

PIJ finds that actors awarding punitive damages in isolation, such as juries, ignore categories. Figure 1, which reproduces a portion of *PIJ*’s Figure 1,³⁸ shows this. This is *PIJ*’s report of the experimental punitive damages awards in the two experimental conditions—first in isolation and then in a context allowing categorical comparisons. The condition labeled “First” on the horizontal axis shows no meaningful difference in median punitive damages awards between “Personal Injury” and “Financial” cases—case category is all but ignored.

Figure 1. Average Median \$ Award

Because jury trials are analogous to the “First” experimental condition, *PIJ*’s experimental data suggest that case categories should have little influence on the level of punitive damages awards, given a level of harm. Therefore, one prediction is that the relation between punitive and compensatory awards will not meaningfully vary across case categories in actual jury cases.

37. *PIJ*, *supra* note 4, at 1175.

38. The figure displayed here is based on the lower left quadrant of *PIJ*’s Figure 1. *PIJ*, *supra* note 4, at 1176 fig.1.

Judge trials are analogous to *PIJ*'s second condition. Over time, judges observe cases from other categories and thus, unlike juries, can determine a punitive award "in the context of a problem from a different category."³⁹ Figure 1's report on the relation between punitive awards and case categories for its second condition generates a second prediction. Since the punitive award in one category of cases (personal injury) more than doubled in relation to the punitive award in another category of cases (financial injury), *PIJ*'s results indicate that case categories should influence punitive award levels in judge-tried cases. At a minimum, case categories' influence should be greater in judge-tried cases than in jury-tried cases.

2. *Increased punitive awards in case categories involving more prominent harm.*

PIJ's results do more than suggest a stronger case-category effect in judge trials than in jury trials. Figure 1 also shows a within-category effect for categories with the more prominent harm. *PIJ* describes this result: "the more prominent harm is assigned . . . a lower dollar value when judged by itself than when directly compared to a harm of a less prominent kind."⁴⁰ Figure 1, using *PIJ*'s data, shows that personal injury punitive awards are higher when a less prominent category of harm—financial injury—is available to place the punitive award in context. So, if one can assign case categories an ordinal harm ranking, *PIJ* predicts that the category perceived as having greater harm will show a positive shift in punitive awards as one moves from jury trials to judge trials, compared to any shift in punitive awards in the category that has less severe harm. Based on *PIJ*'s results, two promising categories for illustrating the greater positive shift are personal injury cases and non-personal injury cases. Personal injury cases should show a greater shift than non-personal injury cases. But *PIJ*'s underlying theory renders it unimportant that any particular category be selected as more or less prominent a harm category. All that is necessary is that one can assign an ordinal ranking of harm to case categories. The higher-ranked category should show a greater shift in punitive awards as one moves from jury to judge trials.

In the context of real-world perceptions about the legal system, *PIJ*'s core prediction is bold. Its theory and experimental results suggest that, in a case category of prominent harm, judges will award higher punitive damages than juries. We suspect that this runs counter to what most observers would guess about the legal system.

39. *Id.* at 1176.

40. *Id.*

C. *The Data*

PIJ thus generates predictions about differential effects of case categories on punitive awards in judge and jury trials, and about the direction of those effects, conditional on identifying categories of greater and lesser harm prominence. To test these predictions, we use data gathered by the National Center for State Courts from 45 of the largest 75 United States counties, under grants from the Bureau of Justice Statistics ("BJS").⁴¹ These data, which come from the National Center's Civil Trial Court Network ("CTCN"), cover one year of judge and jury trials completed in 1996, and provide the least biased sample of state punitive damages awards in state litigation.⁴² The data allow for meaningful comparisons, across case categories, of the relation between punitive and compensatory damages in judge and jury trials.

The CTCN data include all completed trials in 36 counties and a random sample of trials in nine counties. In eight of the sampled counties, well over half the 1996 trials are included in the data. Summary statistics by county are listed in previous work.⁴³ In one county, Harris County, Texas, which includes Houston, 352 judge and jury trials were randomly sampled out of a population of approximately 1,500 trials, with jury trials somewhat more heavily sampled than judge trials. It is the only county in which sampling might distort the actual number of punitive awards in judge and jury trials.⁴⁴ The analyses reported here include Harris County but we report throughout the article whether inclusion of Harris County materially affects results.

Table 1 presents summary statistics of awards studied here. Because *PIJ* reports results for financial injury cases compared to personal injury cases, we divide the data into those cases claiming bodily injury and those cases not claiming bodily injury. Panel A, which shows compensatory awards, shows that the mean compensatory award in cases with punitive damages and a claim for bodily injury was \$555,000 in jury trials and \$1.54 million in judge trials. The median compensatory award for bodily injury cases was \$52,000 in jury trials and \$61,000 in judge trials. Panel B, which shows punitive awards, shows that the mean punitive award for cases with bodily injury claims was \$270,000 in jury trials and \$243,000 in judge trials. The median punitive

41. For a summary of the data and methodology, see CAROL J. DEFANCES & MARIKA F. X. LITRAS, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, CIVIL TRIAL CASES AND VERDICTS IN LARGE COUNTIES (1996), available at <http://www.ojp.usdoj.gov/bjs/pub/pdf/ctcvlc96.pdf>.

42. See Eisenberg et al., *supra* note 17, at 747-48 & n.17, 755 n.40.

43. Eisenberg et al., *supra* note 17, at 780 tbl. 1.

44. In Harris County sampling may have led to an extraordinary proportion of judge-tried punitive damages cases. BJS 1996, *supra* note 41, at 23 (reporting 67 bench trials with punitive awards, which exceeds the sum of bench trial punitive awards in all other counties in the sample). The 67 trials reported reflect the county-level sampling. Twelve actual punitive awards were found in Harris County. Eisenberg et al., *supra* note 17, at 748 n.19.

award for bodily injury cases was \$44,000 in jury trials and \$80,000 in judge trials.

TABLE 1. SUMMARY STATISTICS OF PUNITIVE AND COMPENSATORY AWARDS IN JUDGE AND JURY TRIALS WITH PUNITIVE AWARDS⁴⁵

	MEAN		MEDIAN		NUMBER OF AWARDS	
	Jury	Judge	Jury	Judge	Jury	Judge
A. COMPENSATORY DAMAGES (thousands)						
Bodily injury claimed	\$555	\$1,540	\$52	\$61	36	14
No bodily injury claimed	\$618	\$696	\$98	\$54	83	40
B. PUNITIVE DAMAGES (thousands)						
Bodily injury claimed	\$270	\$243	\$44	\$80	36	14
No bodily injury claimed	\$2,486	\$667	\$57	\$29	83	40

Table 1's award patterns help put this study in perspective. This is not a study of the mass of cases or even of the mass of trials. It includes only those trials, won by plaintiffs, that have both a punitive award and a compensatory award. Such cases are a tiny fraction of completed trials.⁴⁶ The relation between judge and jury awards in this highly filtered sample is not representative of the relation between judge and jury awards in the mass of cases. For example, in the mass of cases mean and median compensatory jury awards are much higher than judge awards.⁴⁷ This is likely a consequence of substantial non-random case routing between judges and juries.⁴⁸ In the

45. Source—BUREAU OF JUSTICE STATISTICS DATA ON 45 LARGE COUNTIES (1996).

46. Eisenberg et al., *supra* note 17, at 749 tbl. 1 (showing 176 punitive awards in over 8000 trials).

47. *Id.*

48. Amounts demanded in jury-tried cases are significantly higher than amounts demanded in judge-tried cases. *Id.* at 766; Kevin M. Clermont & Theodore Eisenberg, *Trial by Jury or Judge: Transcending Empiricism*, 77 CORNELL L. REV. 1124, app. B, at 1177 (1992) [hereinafter Clermont & Eisenberg, *Transcending Empiricism*] (reporting that, in a study of federal court cases from 1979 through 1989, the mean amount demanded in jury trials is larger than in judge trials for all but one category of the cases studied); Theodore Eisenberg & Kevin M. Clermont, *Trial by Jury or Judge: Which is Speedier?*, 79 JUDICATURE 176, 180 (1996) (noting that "[b]oth demands and awards of damages are higher in cases tried to juries than in cases tried to judges"). And tort trials tend to be routed to juries while contract cases tend to be adjudicated by judges. Clermont & Eisenberg, *Transcending Empiricism*, *supra*, tbl.4, at 1141, 1157 (reporting that, in a study of federal court cases from 1979 through 1989, the percentage of judge trials is "uniformly low for personal-injury cases and uniformly high for other cases"); Theodore Eisenberg, John Goerdt, Brian Ostrom & David Rottman, *Litigation Outcomes in State and Federal Courts: A Statistical Portrait*, 19 SEATTLE U. L. REV. 433, 443 tbl.4 (1996) (reporting that, in a study

sample studied here, limited to a group of the most egregious cases, those leading to punitive awards, there is no consistent pattern of jury awards being higher than judge awards. For cases with bodily injury claims, both the mean and the median compensatory awards in judge cases exceed the mean and median compensatory awards in jury cases. Only the median jury award in no-bodily-injury cases exceeds the comparable judge award. The filter limiting the sample to cases with punitive damages substantially influences the characteristics of the sample's compensatory awards.

D. Results

PIJ's results suggest that case categories should play little or no role in explaining punitive award levels in jury-tried cases, but should play some role in judge-tried cases. The results also suggest that judges will produce higher median punitive damages awards in cases that involve more prominent harm. We find some evidence to support these predictions.

PIJ reports median punitive awards, with the compensatory harm fixed in all scenarios at \$500,000.⁴⁹ Real-world data do not supply the luxury of unvarying compensatory awards. We therefore study the ratio between punitive and compensatory awards in each case with a punitive award. *PIJ*'s results forecast that case category will not influence the punitive-compensatory ratio in jury-tried cases and will influence it in judge-tried cases. Furthermore, the ratio will increase most in judge-tried cases involving the more prominent harm category. To implement a case-category dichotomy, we again divide the sample into two categories: cases in which bodily injury was claimed and cases in which no bodily injury was claimed.

Table 2 reports summary statistics for the punitive-compensatory ratio for each of three groups of cases. Panel A reports on the full sample of cases awarding punitive damages. Panel B contains all 44 counties other than Harris County. Panel C limits the sample to the four subcategories of cases that had at least five judge trials and at least five jury trials. It thus reduces the possibly noisy effects generated by judges and juries seeing different categories of cases but still retains over half the original sample. Since *PIJ*'s effects are reported for median awards, Table 2's median columns, its third and fourth numerical columns, are of primary interest.

of both federal and state trials from 1991-1992, the percentage of bench trials is high for contract cases and low for tort cases).

49. *PIJ*, *supra* note 4, at 1174 n.73.

TABLE 2. PUNITIVE-COMPENSATORY RATIO
IN JUDGE AND JURY TRIALS WITH PUNITIVE AWARDS⁵⁰

	MEAN		MEDIAN		NUMBER OF AWARDS	
	Jury	Judge	Jury	Judge	Jury	Judge
A. PUNITIVE-COMPENSATORY RATIO						
Bodily injury claimed	11.13	1.42	.57	1.01	36	14
No bodily injury claimed	1.75	1.16	.47	.53	83	40
B. EXCLUDING HARRIS COUNTY						
Bodily injury claimed	11.75	1.62	.55	.55	34	8
No bodily injury claimed	1.48	1.27	.47	.51	81	34
C. SAMPLE LIMITED TO 4 CASE CATEGORIES WITH 5 OR MORE JUDGE AND JURY TRIALS						
Bodily injury claimed	2.27	1.58	.34	1.44	15	12
No bodily injury claimed	1.11	.64	.36	.53	46	26
D. EXCLUDING HARRIS COUNTY						
Bodily injury claimed	2.27	2.00	.34	1.01	15	6
No bodily injury claimed	1.11	.65	.38	.50	45	21

Panel A supplies some support for *PIJ*'s predictions. The difference between median punitive-compensatory ratios in bodily injury and non-bodily injury cases tried before juries is rather small, falling from .57 in bodily injury cases to .47 in non-bodily injury cases. Case category has little effect on jury-awarded, punitive-compensatory ratios. The difference is substantially larger in cases tried before judges, with the median ratio for bodily injury cases being 1.01—almost double the median ratio of 0.53 in non-bodily injury cases. Case-category thus has a stronger effect on judge-tried cases than jury-tried cases. Furthermore, the effect is in the surprising direction predicted by *PIJ*: The ratio increases in judge-tried cases relative to jury-tried cases. But Panel B suggests that Panel A's results are largely a consequence of including one county—Harris County, Texas. With Harris County removed from the sample, as in Panel B, the predicted effects disappear.

In an effort to stabilize the results, Panels C and D limit the sample to those subcategories of cases that have at least five punitive awards in both judge and jury trials. The four surviving subcategories are premises liability, intentional tort, fraud, and buyer-seller disputes. Panel C, which includes Harris County, shows a strong case-category effect. Median ratios in judge-tried bodily injury cases are four times greater than median ratios in jury-tried bodily injury cases. Panel D again limits the sample to the subcategories of cases with five punitive awards in judge and jury trials. Like Panel B, it excludes Harris County. But

50. Source—BUREAU OF JUSTICE STATISTICS DATA ON 45 LARGE COUNTIES (1996).

here the predicted shift survives the exclusion. Median ratios in judge-tryed cases with bodily injuries are three times greater than median ratios in jury-tryed bodily injury cases. Excluding Harris County reduced but did not eliminate the effect.

To test the significance of the effects, we use a regression model in which the punitive-compensatory ratio is the dependent variable. Multiple regression works to segregate the effects of various factors, such as case category and mode of trial.⁵¹ The explanatory variables are dummy variables for the mode of trial (jury trial or not) and case category (bodily injury or not). We add an interaction term equal to the product of the two dummy variables, jury trial x bodily injury. The interaction term explores the possibility that judges and juries react differently to the presence of bodily injury.⁵² Because we are estimating the median ratio, we use quantile regression rather than ordinary least squares regression, which focuses on the mean.⁵³ Standard errors are calculated using bootstrap resampling with 1000 replications.⁵⁴ Table 3 reports the results.

The models modestly support *PIJ*'s results. The first, third, and fourth models yield results consistent with *PIJ*'s predictions. The coefficient for the interaction term is negative, suggesting lower punitive-compensatory ratios in bodily injury, jury-tryed cases than in judge tryed cases. These three models, however, are not consistent with the second model, which uses all case categories but excludes Harris County. Its coefficients and significance levels differ substantially from the first model, which uses the full sample of 173 cases. But one can reasonably argue that the sample can properly be limited to

51. Multiple regression is a statistical technique that quantifies the influence of each of several factors (independent variables) on the phenomenon being studied (dependent variable). See generally MICHAEL O. FINKELSTEIN & BRUCE LEVIN, STATISTICS FOR LAWYERS 323-28 (1990).

52. For a brief discussion of interaction terms, see JAN KMENTA, ELEMENTS OF ECONOMETRICS 508-10 (2d ed. 1986).

53. Quantile regression estimates the relation between explanatory variables and the median of the dependent variable. Ordinarily least squares regression estimates the relation between explanatory variables and the mean of the dependent variable. For discussion of quantile regression, see Roger Koenker & Gilbert Bassett, *Regression Quantiles*, 46 ECONOMETRICA 33 (1978). For applications of quantile regression, see, e.g., Theodore Eisenberg & Martin T. Wells, *Ranking and Explaining the Scholarly Impact of Law Schools*, 27 LEGAL STUD. 373, 410 (1998); ANGELA HAWKEN, STEPHEN J. CARROLL & ALLAN F. ABRAHAMSE, RAND INST. FOR CIV. JUST., THE EFFECTS OF THIRD-PARTY, BAD FAITH DOCTRINE ON AUTOMOBILE INSURANCE COSTS AND COMPENSATION 39-40 tbls. 3.4 & 3.5 (2001).

54. Traditional statistical procedures make important assumptions, such as sampling from a normal distribution. When such assumptions are not valid, the procedures may fail. Bootstrap resampling offers a more reliable approach that is less dependent on sometimes questionable assumptions. See B. EFRON & R.J. TIBSHIRANI, AN INTRODUCTION TO THE BOOTSTRAP (1993). These data illustrate an important methodological issue. The results of simple quantile regressions, without bootstrap resampling, yield spuriously impressive statistically significant results.

subcategories of cases expected to be adjudicated by both judges and juries. This reduces the likelihood of differences being attributable to differences in the mix of cases adjudicated by judges and juries.

This shifts the focus to the third and fourth models, which are more consistent. The third model, which limits the sample to case subcategories with five or more judge and jury trials in the full sample, and includes Harris County, yields large coefficients, significant at the 0.1 level, for both the bodily injury dummy variable and the interaction term. The interaction term, as noted above, is in the direction forecast by *PIJ*. Jury-tried bodily injury cases have lower median punitive-compensatory ratios than judge-tried bodily injury cases. The bodily injury and interaction term coefficients increase in size in the fourth model, but become statistically insignificant.

TABLE 3. REGRESSION MODELS OF PUNITIVE-COMPENSATORY RATIO⁵⁵

Variable	Dependent Variable = punitive-compensatory ratio			
	Full sample	Full sample without Harris	Four case categories with 5 or more judge & jury trials	Four categories without Harris
Jury trial dummy	-.087 (.558)	-.042 (0.802)	-0.174 (0.198)	-0.116 (0.510)
bodily injury dummy	.863 (.100)	0.090 (.931)	0.897* (0.085)	0.921 (0.503)
Jury trial _ bodily injury	-.719 (.196)	0.054 (0.960)	-0.943* (0.093)	-0.968 (0.490)
constant	.556 (<.0001)	.511 (<.0001)	.556 (<.0001)	.497 (.001)
pseudo r-squared	.0007	.0002	.0295	.0063
number of cases	173	157	99	87

* $P < .1$

Support for *PIJ*'s thesis is most clearly observed in Figure 2. It shows the punitive-compensatory ratio, for both judge and jury trials, for the two categories of harm, "bodily injury" and "non-bodily injury." The data portrayed in the figure are limited to case categories with multiple judge and jury trials—the same samples used in Panel C of Table 2 and in Table 3's third

55. Source—BUREAU OF JUSTICE STATISTICS DATA ON 45 LARGE COUNTIES (1996). Quantile (median) regression of the punitive/compensatory ratio. Standard errors are calculated using bootstrap resampling with 1000 replications. The second and fourth models exclude cases from Harris County, Texas. Significance levels are in parentheses.

regression model. The left endpoint of each line represents the punitive-compensatory ratio in jury trials, the right endpoint of each line represents the punitive-compensatory ratio in judge trials. Because juries see only one case and judges are repeat observers, these endpoints correspond, respectively, to the "First" and "Second" conditions appearing on the horizontal axis in Figure 1. Figure 2 is strikingly similar to Figure 1. The right endpoints are much farther apart than the left endpoints. Judge punitive damages awards in bodily injury cases are high relative to jury awards. Consistent with our numerical analysis, this relation does not emerge in non-bodily injury cases.

Figure 2. Punitive-Compensatory Ratio by Nature of Injury and Adjudicator⁵⁶

Although the results are far from providing conclusive support for *PIJ*'s predictions, the bold prediction of lower awards per unit of compensatory damages in jury-trying bodily injury cases cannot be rejected and has some support. With a larger sample, which is likely to be difficult to gather given the paucity of punitive damage awards, the results might achieve traditional levels of statistical significance. Even so, the direction of the predicted effect is surprising.

E. *Effects of Case Selection*

Interpreting the empirical tests of *PIJ*'s punitive damages thesis requires considering the effect of case selection.⁵⁷ We consider two separable aspects of

56. Source—BUREAU OF JUSTICE STATISTICS DATA ON 45 LARGE COUNTIES (1996).

57. See, e.g., Kevin M. Clermont & Theodore Eisenberg, *Do Case Outcomes Really Reveal Anything About the Legal System? Win Rates and Removal Jurisdiction*, 83 CORNELL L. REV. 581, 588-91 (1998) (describing the "case-selection effect," that is, the proposition

case selection. First, litigants do not randomly route cases to judges and juries. Substantial evidence suggests that high-stakes cases and personal injury cases are routed to juries.⁵⁸ Second, litigants do not randomly settle cases. The characteristics of cases litigated to trial may differ from those that settle.

Perceptions about judge and jury performance should shape case routing between the two adjudicators. People perceive that juries are, relative to judges, more sensitive to emotional content than judges and hence plaintiffs might route more emotionally laden cases to juries than judges.⁵⁹ Although such selection effects could influence our result, this effect would not undermine our support for *PIJ*'s hypothesis. If plaintiffs prefer juries in emotionally laden cases, and the emotional content of a case increases the punitive-compensatory ratio, then our results would understate the support for *PIJ*'s thesis from real cases. If juries disproportionately see emotionally laden cases, then our results overestimate the punitive-compensatory ratio in jury trials, which is already less than the observed ratio in judge trials. Without disparate case routing, the actual judge-trial punitive-compensatory ratio might exceed the jury-trial ratio by an even greater amount than we observed.⁶⁰ We therefore doubt that the routing of cases between judges and juries explains the observed pattern supporting *PIJ*'s theory.

On the other hand, the effects of the settlement process might undermine the support for *PIJ*'s hypothesis that we observe. Because juries are believed to expose defendants to high risk in high-stakes personal injury cases, defendants may choose to settle, rather than try, cases with high punitive award potential.⁶¹ The effects of settlement on case selection have not been conclusively determined.⁶² If jury-tried cases with high punitive potential are

that litigated cases are a non-random sample of underlying disputes, since parties tend to settle, rather than litigate, cases that are clear).

58. Clermont & Eisenberg, *supra* note 48, at 1162-65; Eisenberg et al., *supra* note 16, at 766; Eric Helland & Alexander Tabarrok, *Runaway Judges? Selection Effects and the Jury*, 16 J.L. ECON. & ORG. 306, 324-27 (2000) (finding differences in samples of cases routed to judges and juries); Kelso & Kelso, *supra* note 18, at 21 (hypothesizing a "sampling bias").

59. Clermont & Eisenberg, *Transcending Empiricism*, *supra* note 48, at 1127-28.

60. For further discussion of this issue, see Eisenberg et al., *supra* note 17, at 766-67.

61. See, e.g., David Segal, *Tag-Team Lawyers Make Businesses Blink; HMOs Latest to Grapple With Threat of Investor-Scaring Mega-Verdicts*, WASH. POST, Nov. 12, 1999, at A1 (explaining that litigation's threatened effect on corporate stock prices often spurs companies to settle). Studies funded by industry or tort reform groups, which find evidence that punitive damages often are requested, are not very helpful in addressing this issue. See Eisenberg et al., *supra* note 17, at 767-68.

62. See, e.g., Thomas Koenig, *The Shadow Effect of Punitive Damages on Settlements*, 1998 WIS. L. REV. 169, 170 (providing an overview of the limited research available on the "largely unexplored shadow effect of punitive damages on settlement"); Herbert M. Kritzer & Frances Kahn Zemans, *The Shadow of Punitives: An Unsuccessful Effort to Bring it into View*, 1998 WIS. L. REV. 157, 159-60 (reporting that there is little information available on

settled rather than tried, our case data may not be as supportive of *PIJ*'s thesis as first appears.

Available evidence suggests skepticism about the existence of a "hidden mountain of settlements involving punitive damages masking the ability of trial studies to detect punitive damages' true effects."⁶³ Evidence from the effect on stock prices of settlement of punitive damages claims offers no support for this theory.⁶⁴ In addition, caps on punitive damages reduce the risk of defendants' suffering a catastrophic punitive award. Settlement to avoid a large punitive award should be less prominent in states that cap awards. Yet one cannot detect materially different settlement rates between states that do and do not cap punitive damages.⁶⁵ In another context, efforts to explain away large pro-defendant movements in products liability law based on an unobserved mountain of settlement activity are questionable.⁶⁶

None of this conclusively establishes that fully accounting for settlement would leave real-case support for *PIJ*'s experimental evidence intact. But the evidence that does exist precludes unquestioning reliance on a settlement effect. Given the massive amount of well-funded recent research on punitive damages, it is time for researchers who believe settlement patterns mask important punitive damages effects to present evidence supporting that thesis.

III. IDENTIFYING SOURCES OF COHERENCE AND INCOHERENCE

PIJ's laboratory results reveal a psychological phenomenon that should trouble the legal system. As *PIJ* notes, the apparent arbitrariness with which people translate their qualitative assessments into quantitative ones raises concerns about damage awards, criminal sentences, and administrative penalties.⁶⁷ Our own study provides evidence suggesting that this phenomenon affects the outcome of real cases. The legal system appears to ask people to perform a cognitive task that they cannot accomplish—to translate qualitative judgments into quantitative ones reliably, consistently, and coherently.

The puzzle remains, however. Despite the foundational defect *PIJ* identifies, all credible studies of punitive damages in real cases demonstrate a striking consistency. How can it be that the system relies on individually incoherent judgments and yet produces a coherent whole?

the impact that demands for punitive damages have on either the size of settlements or the costs of handling cases).

63. Eisenberg et al., *supra* note 17, at 768.

64. Karpoff & Lott, *supra* note 20, at 537-39. For reservations about the Karpoff-Lott data, see text accompanying notes 94-101 *infra*.

65. Eisenberg et al., *supra* note 17, at 769-71.

66. Theodore Eisenberg & James A. Henderson, Jr., *Inside the Quiet Revolution in Products Liability*, 39 UCLA L. REV. 731, 744-48, 755-56 (1992).

67. *PIJ*, *supra* note 4, at 1160, 1194-1197.

Exploring this question requires identifying sources of both coherence and incoherence in the legal system. The system contains importance sources of coherence that help explain the sober awards pattern in actual cases. A prime source of incoherence implicated by *PIJ*, the phenomenon of anchoring and adjustment, is a legitimate concern. But we suggest that *PIJ* may have misidentified the anchoring of punitive awards on compensatory awards as evidence of incoherence when the available data suggest it is better viewed as evidence of coherence.

A. Sources of Coherence

To begin to answer how the system produces a coherent whole from individually tainted judgments, it is important to identify what the laboratory studies of cognitive processes show and what they do not show. Such studies reveal how people think about the kinds of problems that legal decisionmakers confront. They can demonstrate, as *PIJ* does, that a critical cognitive facility is lacking, or can demonstrate that particular cognitive processes produce biases in judgment. If the results of such studies are convincing, different conclusions about the legal system are possible: the system is defective in some way and should be fixed; the system has adopted some remedy for the defect; or the system does not actually rely on such judgments to a troubling degree.⁶⁸

The identification of a cognitive limitation of legal decisionmakers in a controlled, experimental study usually leads to the first conclusion—that reform is necessary.⁶⁹ In fact, some of the *PIJ* authors previously have argued that cognitive difficulties in translating qualitative blameworthiness to quantitative punitive monetary damages support greater use of judges or administrative agencies to determine punitive damages.⁷⁰

The difficulty with using *PIJ*'s data to advocate reform is that the laboratory results that it reports do not account for the full range of responses that the legal system already includes that promote coherence. As we have noted, the studies of real cases reveal only a rather limited degree of incoherence. Forces seem to be at work that already reduce incoherence. Thus, the second conclusion, that the system has adopted some remedy (or remedies) for the cognitive limitations of human judgment, may be more appropriate.

68. See Jeffrey J. Rachlinski, *The "New" Law and Psychology: A Reply to Critics, Skeptics, and Cautious Supporters*, 85 CORNELL L. REV. 739, 753 (2000) (suggesting that normative responses to cognitive defects of legal actors depend upon the scope and impact of these defects on legal adjudication).

69. Cf. *id.* at 740-41 (speculating that behavioral studies of the law "should improve both positive and normative legal analysis").

70. See Sunstein et al., *supra* note 2, at 2112-30 (reviewing reform proposals and proposing reforms to rationalize the quantification of damage awards by removing the task from juries).

The legal system, in fact, has had ample opportunity to observe and address the problems that *PIJ*'s results could create. In the case of damage awards, repeat players who process many similar cases, such as large firms and insurance companies, would observe incoherence, if it existed. These repeat players would be able to present evidence of incoherence to make a convincing case for adopting procedures that reduce any observed incoherence. Repeat players within the system, most notably judges, would also observe incoherence in the form of a high variance in awards among cases that present similar damages profiles. Judges would be unwilling to tolerate arbitrary awards and would develop rules that allow them to intervene to reduce the arbitrariness of judgments. Thus, procedural rules may exist precisely to reduce the level of incoherence.

As one example of such a rule, judicial supervision of damage awards arguably reflects a long-standing recognition that individual decisionmaking bodies such as juries might improperly translate their assessments of a case into a dollar amount. Judges have long had the power to review jury awards and overturn those awards that are erratic. Furthermore, appellate judges have the power to supervise trial judges' awards as well as the power to review those judges' review of jury awards. The system of damage awards has thus developed in a way that ensures that experienced decisionmakers review awards.⁷¹

In many cases, the legal system could simply avoid relying on the type of judgment that the experimental laboratory work documents. Laboratory studies rely on problems that are carefully designed to identify the type of thought processes of interest to the experimenters. In contrast, the legal system produces problems that are the product of an adversarial process operating under a set of highly refined rules. In other contexts, courts have avoided relying on judgments that are likely to be affected by undesirable cognitive biases.⁷² For example, rather than judge the potential negligence of medical doctors *ex post*—a judgment likely to be affected by the hindsight bias—courts compare doctors' conduct with standard medical practice.⁷³ Recognizing that judgments made in hindsight are unreliable, courts have shifted their focus.

71. Experienced appellate decisionmakers might be especially sensitive to review of cases with substantial damages awards. For evidence that federal appellate courts tend to reverse plaintiff trial victories more than defendant trial victories, see Kevin M. Clermont & Theodore Eisenberg, *Appeal from Jury or Judge Trial: Defendants' Advantage*, 3 AM. L. & ECON. REV. 125 (2001) (finding that defendants suffer a lower reversal rate of trial victories than plaintiffs); Kevin M. Clermont & Theodore Eisenberg, *Anit-Plaintiff Bias in the Federal Appellate Courts*, 84 JUDICATURE 128 (2000) (same); Kevin M. Clermont & Theodore Eisenberg, *Plaintiphobia in the Appellate Courts: Civil Rights Really Do Differ from Negotiable Instruments*, 2002 U. ILL. L. REV. (forthcoming 2002) (on file with author) (same results in fuller sample).

72. See Rachlinski, *Judging in Hindsight*, *supra* note 14, at 606-08.

73. *Id.* at 612.

Furthermore, the adversarial nature of most legal decisionmaking processes creates another opportunity for coherence. In laboratory research on cognitive biases, the experimental materials remain constant (with only controlled manipulations); but, of course, in the real world, the objects of judgment are pliable. Typically, only one of the two adversaries in a lawsuit will benefit from a cognitive bias, and thus, the other one will try to restructure the evidence so as to reduce the influence of the bias. For example, the hindsight bias seems to benefit tort plaintiffs by making defendants seem like they should have been able to foresee, and therefore avoid, injuring the plaintiff.⁷⁴ But a nimble defendant can turn the bias against the plaintiff by arguing that the plaintiff could also have foreseen the injury and should therefore have avoided the circumstances that led to the injury.⁷⁵

Furthermore, the settlement process creates a powerful source of coherence. Across almost every area of law, parties settle the vast majority of cases that are not decided by a dispositive motion.⁷⁶ The settlement process is managed largely by attorneys, most of whom will have some experience in the area of law in which the case falls, or can gain access to other attorneys with such experience. Consequently, these attorneys will understand the range of awards a case is likely to bring, which in turn will affect their settlement discussions. Although plaintiffs' attorneys may want to gamble that a jury will adopt an unusually high conversion factor when deciding on the appropriate damage award, we doubt it. Plaintiffs have many incentives to make risk-averse settlement decisions, and such gambles are apt to be rare.⁷⁷ Although cognitive defects in judgment undoubtedly affect settlement decisions,⁷⁸ the problem that *PIJ* identifies is largely avoided by the settlement decisions of repeat players who understand the kinds of awards the legal system is apt to produce.

Thus, it is unclear whether *PIJ*'s findings have normative implications for reform, have positive implications for assessing the forces that created the legal system that we now observe, or identify pitfalls for litigants to avoid. In any

74. *See id.* at 600.

75. *See id.* at 594-95.

76. *See* Marc Galanter, *Real World Torts: An Antidote to Anecdote*, 55 MD. L. REV. 1093, 1100 n.17 (1996) (estimating that 7 percent of cases end in trials, 24 percent are decided on dispositive motions, and the rest settle).

77. *See* Jeffrey J. Rachlinski, *Gains, Losses, and the Psychology of Litigation*, 70 S. CAL. L. REV. 113, 119-20 (1996) (reviewing evidence that plaintiffs make risk-averse settlement decisions).

78. *See id.*; *see also* Chris Guthrie, *Framing Frivolous Litigation: A Psychological Theory*, 67 U. CHI. L. REV. 163, 185-95 (2000) (discussing how settlement negotiations are affected by cognitive perceptions of the dispute rather than neutral risk and cost-benefit assessments); Russell Korobkin & Chris Guthrie, *Psychological Barriers to Litigation Settlement: An Experimental Approach*, 93 MICH. L. REV. 107, 128-42 (1994) (same); George Loewenstein, Samuel Issacharoff, Colin Camerer & Linda Babcock, *Self-Serving Assessments of Fairness and Pretrial Bargaining*, 22 J. LEGAL STUD. 135 (1993).

case, numerous factors create a degree of coherence in real cases that *PIJ*'s authors would not have observed in their study. Despite the defect in cognitive judgment that *PIJ* identifies, the system can still produce coherent, reasonable awards—which is what one observes.

B. *Sources of Incoherence: Anchoring as a Case-in-Point*

Regardless of the legal system's coherence-promoting responses, *PIJ* remains important because the system doubtlessly retains some of the incoherence that *PIJ* identifies. Indeed, despite the well-documented regularity of the mass of awards, we do observe the effect that *PIJ* has predicted in a set of cases. The mechanisms that correct for incoherence seem to allow some deficiencies in human judgment to manifest themselves.

The problem of anchoring illustrates how the system can fail to correct itself, and anchoring plays a key role in *PIJ*'s indictment of the current system. Anchoring refers to the tendency, when making numeric estimates, to rely on fixed reference points (or anchors) and to adjust estimates based on these fixed reference points.⁷⁹ If the translation of qualitative categories to quantitative assessments is as difficult as *PIJ* suggests, then anchoring surely plays havoc with the damage awards that litigants make. Even an irrelevant anchor can alter a decisionmaker's thought process or provide the modulus that *PIJ* indicates is necessary to make quantitative legal judgments.⁸⁰

Psychologists attribute anchoring to two cognitive processes.⁸¹ First, anchors create mnemonic associations that affect the final estimate. As part of evaluating the anchor, people mentally test its accuracy with examples or theories that then become salient. These examples or theories, in turn, affect the final estimate.⁸² Second, simple cognitive laziness can be at work. Adjusting from an anchor requires people to consider reasons why the actual

79. See Tversky & Kahneman, *supra* note 24, at 1128-30.

80. See SCOTT PLOUS, *THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING* 146 (1993) (citing an unpublished study by George Quattrone and colleagues).

81. See Timothy D. Wilson, Christopher E. Houston, Kathryn M. Etling & Nancy Brekke, *A New Look at Anchoring Effects: Basic Anchoring and Its Antecedents*, 125 J. EXPERIMENTAL PSYCHOL.: GEN. 387, 388 (1996) (describing two anchoring theories in which (1) a person rationalizes the significance of the anchor value by focusing on that value's putative relevance to the sought-after target value; and (2) a person merely accepts unquestioningly the relevance of some arbitrary anchor value).

82. See Gretchen B. Chapman & Eric J. Johnson, *Anchoring, Activation, and the Construction of Values*, 79 ORG'L BEHAV. & HUM. DECISION PROCESSES 115, 118-19 (1999); Fritz Strack & Thomas Mussweiler, *Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility*, 73 J. PERSONALITY & SOC. PSYCHOL. 437, 438 (1997).

estimate is higher or lower than the anchor, which might require more mental effort than people are willing to commit to the task.⁸³

Anchoring creates two problems for the legal system: Irrelevant anchors influence people's judgment, and people fail to make adequate adjustments from relevant, but extreme, anchors. The legal system can introduce anchors that are arguably irrelevant to the estimation task. For example, one recent experiment showed that when federal magistrate judges first ruled on a frivolous motion to dismiss a tort suit for failing to meet the jurisdictional amount in federal court for a diversity case (\$75,000),⁸⁴ judges gave damage award estimates that were about thirty percent smaller than when they had not ruled on such a motion.⁸⁵ The motion to dismiss apparently served as a low anchor on awards.

The system can also introduce misleading anchors, if adjustment is insufficient. For example, several mock-jury studies show that damage-award requests made by plaintiffs' attorneys have a large impact on jurors' damage awards—even when such requests are outrageously high.⁸⁶ Other anchors, such as statutory damage caps, insurance policy limits, and settlement talks, all can provide numeric anchors that can then influence damage awards. To be sure, anchors can be a source of coherence, if relevant anchors are purposely integrated into the system. Damage caps that are deliberately meant to scale the awards,⁸⁷ narrowly tailored sentencing guidelines, and statutory multipliers (such as in antitrust cases)⁸⁸ could all provide deliberately considered anchors that promote coherence.

More often than not, however, the legal system introduces arbitrary anchors that lead the process astray. Statutory damage caps are most often the result of a political victory by an interest group, and are thus designed to reduce the amount of the awards, rather than promote coherent damage awards. Similarly, sentencing guidelines could reflect an effort to increase penalties to be tough on crime, rather than an attempt to create uniformity. Also, often anchors are asymmetrically introduced. Defendants often refuse to suggest

83. See Nicolas Epley & Thomas Gilovich, *Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors*, 12 PSYCHOL. SCI. 391, 395 (2001) (noting that, "although people try to adjust their impressions to accommodate situational influences [such as anchors], they adjust too little").

84. See 28 U.S.C. § 1332 (federal diversity jurisdiction statute).

85. See Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 792 (2001).

86. Gretchen B. Chapman & Brian H. Bornstein, *The More You Ask for the More You Get: Anchoring in Personal Injury Verdicts*, 10 APP. COG. PSYCHOL. 519, 526-27 (1996).

87. But damages caps that do not vary across case categories would seem to promote the category-bound incoherence that *PIJ* detects. See Part IV.B *infra*.

88. See Clayton Act § 4, 15 U.S.C. § 15 (providing treble damages for private antitrust plaintiffs).

damage awards for fear that doing so would undermine their ability to simultaneously argue that they are not liable to the plaintiff.

Anchoring thus illustrates how the legal system might fail to correct for decisionmakers' cognitive limitations, thereby creating incoherent judgments. Sometimes lawmakers fail to appreciate the limitations of human judgment. For example, politically motivated statutory damage caps are intended to truncate awards, but introduce an anchor that might unintentionally influence decisionmakers. Other times, the nature of the adversarial process allows one side to take advantage of a cognitive limitation. For example, plaintiffs can make damage requests that anchor decisionmakers, leaving defendants the difficult choice of whether to provide their own (lower) anchoring, which in turn might undermine their efforts to convince a decisionmaker that they are not liable.

C. *Distinguishing Coherence from Incoherence*

The anchoring problem is an important instance of the need to distinguish coherence from incoherence carefully. In the case of punitive damage awards, *PIJ* views the strong relation between punitive and compensatory damages as a troublesome anchoring effect that ought to be reduced.⁸⁹ "We hypothesize that the comparative situation may alter judgments in another way, by reducing the anchoring effects of compensatory damages on punitive awards."⁹⁰ The adjudicator with a comparison base will award punitive damages in a manner that is less related to compensatory damages. *PIJ* seems to regard that reduced relation as a step towards coherence.

The assertion that compensatory awards have an undesirable anchoring effect on punitive awards is not self-evident. We doubt that the relation between punitive and compensatory awards ought to be viewed as a troublesome anchoring effect. As *PIJ*'s Equation (1) suggests, a well-established relation exists between punitive and compensatory awards. Defendants who do more harm should generally be punished more, and compensatory awards are the best available measure of harm caused. The amount of harm that a defendant causes will surely correlate with the outrageousness of their conduct and hence should correlate with the size of a punitive damage award. Any influence compensatory awards have on punitive awards seems like a validation more of the theory underlying Equation (1) than of any incoherence arising from anchoring. Compensatory awards do not create the first type of anchoring problem—reliance on an inappropriate anchor.

PIJ must therefore be referring only to the second type of anchoring problem—insufficient adjustment from an anchor. On this theory, it is possible

89. See *PIJ*, *supra* note 4, at 1178.

90. *Id.* at 1178 & n.80.

that juries are more prone to inadequate adjustment than are less category-bound decisionmakers. Nevertheless, three difficulties with *PIJ*'s treatment of anchoring effects arise: (1) the quality of evidence offered to support the effect's existence, (2) the presence of the same effect in judge-tried cases, and (3) the impossible dilemma that *PIJ*'s critique of anchoring poses. Together, these concerns cast doubt on whether incoherence-producing anchoring is at work.

First, *PIJ*'s evidence that anchoring distorts award patterns in financial cases more than in cases of physical injury is not persuasive. *PIJ* describes this concern immediately after it raises the question of anchoring. The next sentence suggests why *PIJ*'s authors are concerned about an anchoring effect. It states, "In the real world of punitive damages, unlike our experiment, compensatory awards are generally much larger in financial cases than in cases of physical injury."⁹¹ *PIJ* supports this statement with a study by Karpoff and Lott described as "showing mean awards of \$14.8 million in fraud cases, and \$20.6 million in business negligence cases, but \$6.2 million in product liability cases, \$1.6 million in malpractice cases, and \$991,000 in motor vehicle accident cases."⁹² *PIJ* completes this line of thought by stating that in "this pair of categories [financial cases and physical injury cases], the anchoring effects favor large judgments for the less prominent harm."⁹³ Physical injuries involve qualitatively worse behavior, but financial injury cases receive larger quantitative awards.

But the authors may be slaying a nonexistent anchoring dragon. Their evidence of troublesome compensatory anchoring effects in financial cases is from the Karpoff-Lott data. The Karpoff-Lott data set contains results found in no other data set covering civil litigation.⁹⁴ Karpoff and Lott studied twelve years (1985 to 1996) of lawsuits in which plaintiffs sought punitive damages from publicly traded companies.⁹⁵ Their data show ninety-four settlements and 1595 plaintiff verdicts⁹⁶—a settlement-verdict ratio suggesting that the vast majority of plaintiff recoveries are via verdicts rather than settlements. This result radically contradicts virtually every other study of civil litigation. Their finding that 310 of 352 products liability verdicts were for plaintiffs contradicts all other studies of products liability litigation, as does their report that, across all lawsuit types, plaintiff verdicts were eighty-five percent of all verdicts.⁹⁷

The CTCN data, unlike the Karpoff-Lott sample, provide little evidence of a troublesome anchoring effect in financial cases. In the CTCN data, the mean

91. *Id.* at 1178.

92. *Id.* at 1178 n.81.

93. *Id.* at 1178.

94. Eisenberg et al., *supra* note 17, at 768 n. 91.

95. Karpoff & Lott, *supra* note 20, at 527, 538-39.

96. *Id.* tbl. 2.

97. *E.g.*, Eisenberg et al., *supra* note 11, at 635 tbl. 1.

compensatory award in bodily injury cases is \$831,000 compared to a mean of \$643,000 in non-bodily injury cases. Other data sets are also more consistent with the CTCN data than with the Karpoff-Lott data. In federal trials terminating from 1978 to 2000, data from the Administrative Office of the United States Courts show a mean award in personal injury tort cases, adjusted for inflation to year-2000 dollars, of \$1.36 million.⁹⁸ The comparable mean for tried contract cases was \$1.35 million. The median tort award was \$161,000 compared to a median contract award of \$116,000.⁹⁹ Federal cases thus have larger stakes than state cases but neither class of cases supports the Karpoff-Lott result on which *PIJ* relies.

CTCN's 1992 data also show no striking excess of awards in financial injury cases compared to personal injury cases.¹⁰⁰ About eight percent of the awards in the 1992 study were \$1 million or more. The largest proportion of these were in two personal injury case categories, medical malpractice (twenty-five percent of awards were over \$1 million) and products liability (fifteen percent of awards were over \$1 million).¹⁰¹ Both the CTCN data, which account for a substantial fraction of state court litigation, and the federal data, which account for all nonbankruptcy federal civil cases, suggest that the Karpoff-Lott data are anomalous. There may not be any financial case anchoring effect of concern.

Second, if compensatory awards generate an anchoring effect on punitive awards by juries, evidence exists that it exerts a substantially similar anchoring effect on punitive awards by judges. The CTCN data analyzed above show that the same basic relation between punitive and compensatory damages exists for the presumably less category-bound judges as for the more category-bound juries.¹⁰² And this similarity emerges notwithstanding the mild evidence we present above in support of *PIJ*'s category-boundedness argument. Category-boundedness may well affect juries more than judges. Yet, despite this distorting effect, the very same relation between punitive and compensatory awards emerges in both judge and jury trials.

Third, to the extent a relation exists between compensatory and punitive awards, treating that relation as an undesirable form of anchoring leaves decisionmakers (judges or juries) in an impossible situation. Decisionmakers who are somehow shielded from the knowledge of the compensatory award (perhaps through bifurcated proceedings with separate juries) would likely produce awards that are *less* coherent than decisionmakers operating under the

98. Theodore Eisenberg & Kevin M. Clermont, *Judicial Statistical Inquiry Form*, available at <http://teddy.law.cornell.edu:8090/questata.htm>. (last visited Dec. 16, 2001).

99. *Id.*

100. See Bureau of Justice Statistics Special Report: Civil Justice Survey of State Courts, 1992: Civil Jury Cases and Verdicts in Large Counties 5 (1995) (Table 6), available at <http://www.ojp.usdoj.gov/bjs/pub/pdf/cjcavilc.pdf>.

101. *Id.* at 6.

102. Eisenberg et al., *supra* note 17, at 752-55, 772-74.

conventional system.¹⁰³ To criticize jury punitive awards on the ground that they *do* relate to compensatory awards leaves the legal system with no acceptable treatment of the punitive-compensatory relation.

IV. IMPLICATIONS OF *PIJ*'S FINDINGS

PIJ includes a candid discussion of its findings' implications for alternatives to current punitive damages systems. To *PIJ*'s authors' credit, they recognize that greater judicial responsibility over punitive awards is not a panacea.¹⁰⁴ They recognize that anchoring is a powerful enough phenomenon to affect judges as well as juries.¹⁰⁵ Statutory provision of meaningful anchors (such as suggested multipliers or divisors) may well be the only sensible cure for anchoring.

Nevertheless, *PIJ* claims its incoherence evidence provides reason "to favor judicial decisions over jury decisions" and, with appropriate hedging, calls for a shift in power from juries to judges.¹⁰⁶ But *PIJ* is not committed to this as necessarily the most desirable reform. After reviewing the limitations of a mere shift towards increased judicial responsibility, *PIJ* discusses an agency-like bureaucratic response, in which bureaucrats administer "a kind of 'grid' of penalties, matched to the individual facts of diverse cases."¹⁰⁷

We first address whether *PIJ* demonstrates enough incoherence to warrant reform. We then note that the shift towards bureaucracy requires reforming a good deal more than the punitive damages system. Lastly, we consider a concrete reform suggested by *PIJ*'s findings—the need for increased punitive damages awards by juries.

A. *The Degree of Incoherence and the Bureaucratic Shift to Bureaucratic Incoherence*

PIJ's discussion of specific reforms stands against a background of skepticism about reform. The "ideal of seeking a system that is both coherent and reflective of popular moral sentiments can be questioned on both

103. See Stephan Landsman, Shari Diamond, Linda Dimitropoulos & Michael J. Saks, *Be Careful What You Wish For: the Paradoxical Effects of Bifurcating Claims for Punitive Damages*, 1998 WIS. L. REV. 297.

104. *PIJ*, *supra* note 4, at 1182-1183. Anecdotally, even high-ranking judges seem capable of misunderstanding the relation between punitive and compensatory awards. Theodore Eisenberg, *Damage Awards in Perspective: Behind the Headline Grabbing Awards in Exxon Valdez and Engle*, 36 WAKE FOREST L. REV. 1129, 1142-45 (2001) (suggesting that the Ninth Circuit in *Exxon Valdez* misunderstood the relationship between the punitive and compensatory award).

105. See Guthrie et al., *supra* note 85 (demonstrating that anchoring affects judges).

106. *PIJ*, *supra* note 4, at 1156.

107. *Id.* at 1184.

normative and practical grounds.”¹⁰⁸ We think there is reason for empirically based skepticism as well. *PIJ* offers reform, admittedly imperfect, because it regards the existing state of affairs as scandalous. “[W]e submit that incoherence, once acknowledged, is unacceptable, even scandalous.”¹⁰⁹ “Scandalous” is a strong term and depends, we submit, on finding not merely evidence of incoherence, but evidence of very substantial incoherence. We do not think that *PIJ*’s data or our own meet that standard.

The effects *PIJ* finds are likely substantially smaller in the real world than in the experimental world. The structure of *PIJ*’s punitive damages scenario likely elicited punitive damages awards that far exceed what would be observed in actual practice. Those systematically higher awards inflate the category-boundedness effect.

First, evidence suggests that the experimental awards are unusually high. *PIJ* reports in its Appendix B that the compensatory award for its personal injury and financial injury scenarios was \$500,000. Figure 1, from *PIJ*, shows median punitive awards ranging from about \$700,000 to \$2.25 million.¹¹⁰ Since the compensatory award is fixed at \$500,000 in the experiment’s cases,¹¹¹ *PIJ*’s range of reported punitive-compensatory ratios is from about 1.4 to 4.5. The CTCN real-world data indicate that, for compensatory awards of \$500,000 or more, the median punitive-compensatory ratio is 0.46. This ratio is not highly sensitive to the \$500,000 cutoff point; it falls only to 0.44 for compensatory awards of \$100,000 or more. Thus, *PIJ*’s hypothetical cases generate punitive-compensatory award ratios that range from about three times to almost ten times actual case data. These substantially higher ratios suggest that *PIJ*’s observed category-based shift exceeds what one would observe in the real world.

Second, *PIJ*’s scenarios share a distorting feature with other recent punitive damages studies. *PIJ*’s scenarios may elicit greater punitive damages awards from jurors because the experimenters offer jurors only punitive damages as a remedy. The compensatory award of \$500,000 is a given. A punitive award is the only way for the jurors to express disapproval of the defendant’s behavior.¹¹² This may well generate higher and more frequent punitive awards than would exist in real cases where jurors can express disapproval through a compensatory award, and rarely find it necessary to express further disapproval through a punitive award.

108. *Id.* at 1181.

109. *Id.*

110. The \$700,000 figure comes from Figure 1; the \$2.25 million figure is reported at *PIJ*, *supra* note 4, at 1177.

111. *PIJ*, *supra* note 4, at 1174 n.73.

112. Similar concerns exist about other experiments casting doubt on jurors’ punitive damages decisionmaking. See Eisenberg et al., *supra* note 17, at 765 (discussing Reid Hastie & W. Kip Viscusi, *What Juries Can’t Do Well: The Jury’s Performance as a Risk Manager*, 40 ARIZ. L. REV. 901, 905 (1998)).

PIJ's inflated category-boundedness effect should be considered together with real-world evidence of coherence. As shown elsewhere, a surprising coherence emerges in the pattern of punitive damages awards, and this coherence survives across categories.¹¹³ Punitive awards relate to compensatory awards in a way that suggests jurors are punishing most harshly those who cause the most harm. The relation does not depend on case categories and the judiciary already plays a substantial role in assuring that this coherence obtains.¹¹⁴ This relation is not evidence of coherence only if the anchoring effect that concerns *PIJ*, in which compensatory awards unjustifiably anchor punitive awards, is at work. We have already shown that this is not likely.

Even if incoherence of greater magnitude were observed, *PIJ*'s alternative solution—a shift to bureaucracy—cannot improve punitive awards' coherence on its own. For *PIJ* presents substantial evidence of incoherence in the administrative state.¹¹⁵ To achieve greater punitive damages coherence through a bureaucratic shift, one would have to reform punitive damages by shifting them to administrative jurisdiction. Then one would have to assure that the administering bureaucracy did not merely substitute the existing form of bureaucratic incoherence for the incoherence detected in traditional adjudication.

B. *Promoting Coherence Within the Current System*

If *PIJ* is correct, an important implication is that punitive awards in jury-tried personal-injury cases can be too low.¹¹⁶ *PIJ* prefers decisionmakers who are not category-bound. Only the non-category-bound decisionmaker produces punitive awards in personal injury cases that maintain a proper gap in relation to punitive awards in financial injury cases. As Figure 1 illustrates, that gap is largely a consequence of artificially low punitive awards by the category-bound decisionmaker in personal injury cases. A simpler reform than those recommended by *PIJ* would be to systematically increase jury punitive awards in certain classes of cases involving prominent harm, as is done by trebling awards in antitrust cases.

That, of course, is extremely unlikely. More realistically, *PIJ* supports refraining from, or even repealing, reforms that promote incoherence. In particular, *PIJ* suggests that arbitrary caps on punitive awards are ill-advised, at least at current levels. If *PIJ* is correct, then artificial caps on punitive awards increase, not decrease, the category-boundedness problem. For they restrain the

113. See, e.g., Eisenberg et al., *supra* note 17; authorities cited *supra* note 20.

114. See Eisenberg & Wells, *Predictability on Appeal*, *supra* note 20, at 80-83.

115. *PIJ*, *supra* note 4, at 1190 tbl. 2.

116. *PIJ* recognizes that its finding may call for increases in punitive awards. *Id.* at 1181.

decisionmaker from awards that may be needed to achieve cross-category coherence. If bodily injury cases demand a multiplier compared to financial injury cases, that solution cannot readily be achieved within the context of existing capping systems, which often cap punitive awards at two or three times the compensatory award without regard to case category.

V. CONCLUSION

PIJ contributes to legal scholarship on two levels. Methodologically, *PIJ* presents an opportunity to test sophisticated social science theory and experimental findings with real-world data. Law and economics, arguably the dominant social science methodology in law schools, has tended to be long on theory and short on empirically testable hypotheses.¹¹⁷ *PIJ*'s psychological approach probably is not inherently more amenable to empirical testing than other approaches, but theorists in other fields, including law and economics, would do well to emulate *PIJ*'s careful formulation of testable hypotheses.¹¹⁸

More substantively, *PIJ*'s experimental results, combined with substantial real-world coherence, raise the issue of how much coherence is enough for a system to function well. Given the cognitive difficulty of translating qualitative judgments into the quantitative ones, it is worth noting the dearth of evidence that the system of punitive damages has seriously undermined the tort system or the economic system that it helps regulate.¹¹⁹ The data from real

117. See generally Robert C. Ellickson, *The Case for Coase and Against "Coaseanism,"* 99 YALE L.J. 611 (1989) (discussing this phenomenon in studies of transaction costs); see also Peter H. Schuck, *Why Don't Law Professors Do More Empirical Research?*, 39 J. LEGAL EDUC. 323 (1989) (alleging that the lack of empirical research in legal scholarship is harmful to the profession).

118. *PIJ*'s authors reject a prominent law and economics approach to punitive damages, in which optimal deterrence drives punishment. A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 870-76 (1998). See also *Perez v. Z Frank Oldsmobile, Inc.*, 223 F.3d 617, 621 (7th Cir. 2000) (Easterbrook, J.). *PIJ* states that this approach is "in conflict with ordinary intuitions." *PIJ*, *supra* note 4, at 1155 n. 6.

Because of the retributive nature of the underlying intuitions, the public sense of what is just punishment is radically at odds with the idea, popular in some academic circles, that punishment should be grounded largely in its deterrent function, and therefore directly linked to the likelihood that an act will be discovered and punished.

PIJ, *supra* note 4, at 1167 (footnote omitted). Thus, both fellow tort-reform advocates and the Supreme Court reject a narrowly focused deterrence approach to punitive damages. See *Cooper Indus., Inc. v. Leatherman Tool Group, Inc.*, 532 U.S. 424, 438 (2001).

119. An ambitious effort to demonstrate widespread economic costs of the punitive damages system is unpersuasive. Professor Viscusi sought to demonstrate the absence of interstate differences in accident deterrence based on whether states allow punitive damages. W. Kip Viscusi, *The Social Costs of Punitive Damages Against Corporations in Environmental and Safety Torts*, 87 GEO. L.J. 285 (1998). The study has serious flaws. Eisenberg et al., *supra* note 17, at 758 n.49; Eisenberg & Wells, *Predictability on Appeal*,

cases reveal a punitive damages system that appears to be ordered about as well as the system of administrative penalties that *PIJ* criticizes. This occurs despite the reliance on inexperienced, category-bounded decisionmakers for the assessment of punitive damages.

Would-be reformers relying on cognitive limitations must also recognize that reforms are costly, and can introduce unintended consequences. Even supposing, for example, that compensatory awards inappropriately influence punitive awards, a likely cure, bifurcated jury trials, might be worse than the disease. Bifurcation would make the process more expensive and may not have the desired effect. And, unable to assign punitive awards, the first jury might be inclined to let efforts to punish the defendant leak into their assessment of the compensatory awards (or the second jury might be likely to let compensatory issues leak into their assessment of the punitive award). Research on mock juries shows that even though juries can misuse information (such as anchors), juries generally perform much better if they are better informed as to the consequences of their actions.¹²⁰

In the final analysis, it might be best to live with some variability, to try to work marginal improvements in the system, but to acknowledge the real-world data suggesting that the current system is essentially reasonable. Academic and policymaker energy might be best devoted to other societal problems.

supra note 20, at 73 n.45, and is premised on the discredited pure deterrence rationale for punitive damages. See note 118 *supra*.

120. Shari Seidman Diamond & Jonathan D. Casper, *Blindfolding the Jury to Verdict Consequences: Damages, Experts, and the Civil Jury*, 26 LAW & SOC'Y REV. 513 (1992).