

July 28, 2004

**Experimental Simulations and Tort Reform:
Avoidance, Error and Overreaching in
Sunstein et al.'s *Punitive Damages* (2002)***

**Neil Vidmar
Duke Law School**

This article is part of the 2004 Randolph W. Thrower Symposium: The Future of Tort Reform: Reforming The Remedy, Re-balancing the Scales, Emory School of Law, February 19, 2004. It will be published in edited form in Emory Law Journal.

* This research was supported, in part, by a Eugene T. Bost Research Leave from Duke Law School. I am indebted to the following colleagues for comments and other assistance on an earlier draft of the paper: Neal Feigenson, Edie Greene, Valerie Hans, Richard Lempert, Kara MacKillop, Mary Rose, Horace Robertson, Michael Saks and Eulena Small. I invited each of the authors of *Punitive Damages* to respond to an earlier version of this article. Professor Viscusi did respond and offered an important correction regarding my reading of one of his previous writings. The opinions in the article are solely my own and do not necessarily reflect the opinions of Duke Law School nor the opinions of my colleagues.

**Experimental Simulations and Tort Reform:
Avoidance, Error and Overreaching in Sunstein et al.'s *Punitive Damages* (2002)**

Neil Vidmar

Duke Law School

“The present empirical studies of all stages of the punitive damages decision show that the major locus of unreliability and disorder in punitive damages decisions is in jurors’ assessments of an appropriate dollar award.”

Sunstein et al., *Punitive Damages: How Jurors Decide* (2002)

“When you come to a word or sentence you do not understand, or that is unfamiliar to you, it is not permissible to just say ‘grasshopper’ and jump over it. You are in danger of losing the meaning of what you are reading.”

Elementary school teacher, Maple Street School, Gillespie, Illinois, *circa* 1950

I. Introduction

This article addresses tort reform claims made in Cass R. Sunstein, Reid Hastie, John W. Payne, David A. Schkade and W. Kip Viscusi in *Punitive Damages: How Juries Decide*¹ and related articles, research that was largely underwritten by the Exxon Corporation.² Based upon a series of simulation experiments, those authors have made a general claim that juries are incapable of making coherent judgments about punitive damages. In this article I raise serious methodological problems

¹ Cass Sunstein et al. PUNITIVE DAMAGES: HOW JURIES DECIDE (2002) (hereinafter “Punitive Damages”).

² See Amon, Exxon Bankrolls Critics of Punitives, Then It Cites the Research in Appeal of \$5.3 Billion Valdez Award, NATIONAL LAW JOURNAL, May 17, 1999; Alan Zarembo, Funding Studies to Suit the Need; In the 1990s, Exxon Began Paying for Research into Juries and the Damages They award. The Findings Have Served the Firm Well in Court, LOS ANGELES TIMES, December 3, 2003; William R. Freudenburg, The Intersection of Corporate Cash, Science and the Law: Toward a Closer Examination, Paper Abstract, Session on Law and Society: Legal Institutions and Processes, Annual Meeting of the American Sociological Association, August 16, 2003. I raise here, in truth not very subtly, the role of Exxon in funding the studies. However, I leave for another day, and perhaps for other writers, the issue of the role of Exxon in vetting which studies would be funded and which not, the use of the studies in Exxon’s own litigation, the fact that most of the experiments, albeit not all, were published in student-edited law reviews rather than in peer-reviewed social science journals, and other matters. These issues deserve more extensive treatment, see, e.g. see Richard Lempert, Juries, Hindsight Bias, and Punitive Damage Awards: Failures of a Social Science Case for Change, 48 DEPAUL LAW REVIEW 867, 868-71, including note 16 (1999). However, in the remainder of this article I limit myself to addressing on their own merits the research experiments and the authors’ conclusions from those experiments.

bearing on the validity of the research, and, therefore, its ability to provide judges and legislators with useful information about juries and punitive damages.

The principal conclusion about juries is stated in the concluding chapter of *Punitive Damages*:

The many systematic patterns of behavior that we observed are convincing evidence of the jurors' conscientiousness. Nonetheless, the legally required decision tasks often seemed to exceed their individual and social capacities. The decision task is not well defined by jury instructions; jurors are not provided with the necessary background information or experiences to make reliable judgments; and inherent cognitive limitations interfered with their performance of the specific judgments prescribed by our punitive damages system. Jurors' good intentions and high levels of motivation were thwarted by the inherent complexity of the legal decision task and by the lack of clear instructions or other effective guidance. The result is a decision process that is unreliable, erratic and unpredictable.³

Based on the above conclusions, the book makes a policy recommendation that would remove punitive damages decisions entirely from the jury: "...serious consideration should be given to moving away from the jury and toward a system of civil fines, perhaps through a damages schedule of the sort that has been used in many areas of the law, including workers' compensation and environmental violations."⁴

These claims take place at a time and context in which the Supreme Court and lower courts have struggled with developing legal doctrine to address outlier awards involving punitive damages.⁵ Simultaneously, state and federal legislators are considering reforms to the American tort system that also involve punitive damages. The authors of *Punitive Damages* are all eminent scholars in their fields and, combined with the problem of outlier awards, their conclusions have gained the attention of courts and policy makers. *Punitive Damages* and articles upon which the book draws have been cited by the Exxon Corporation in its own litigation⁶ and in amicus briefs before the Supreme

³ PUNITIVE DAMAGES at 241. To be fair, in the final two paragraphs of the book Cass Sunstein cautiously states that "[o]ur findings do not point directly to social reforms, but in the next paragraph they further conclude that, "[i]f our findings are correct, perhaps the ideal system of punitive damages would not involve juries or even judges, but specialists in the subject matter at hand, who would be able to create clear guidelines for punitive awards," at 258.

⁴ PUNITIVE DAMAGES at 242; see also 257.

⁵ E.g. *State Farm Mutual Automobile Insurance Company v. Campbell*, 538 U.S. 408 (2003); *Cooper Industries, Inc v. Leatherman Tool Group, Inc.*, 532 U.S. 424 (2001); *BMW of North America v. Gore*, 517 U.S. 559 (1996);

⁶ *In Re The Exxon Corp. et al.*, U.S. Ct. App., 9th Cir. No. 97-35191 (1997); American Tort Reform Association, Amicus Brief in Support of Appellants Exxon Corporation et.al, in *In Re The Exxon Corp. et al.*, U.S. Ct. App., 9th Cir. No. 97-35191 (1997).

Court.⁷ *Punitive Damages* or the original articles upon which it is based, have been cited in dictum by the Supreme Court.⁸ Some lower courts have recently treated the book as social science authority. In *In re Simon II Litigation*, for example, Judge Weinstein stated:

...in a pathbreaking empirical multidisciplinary study, *How Juries Decide*, the authors demonstrated that, while jurors can agree on the degree of moral and ethical features of defendants who cause massive harm, conscientious as they are, they have no criteria or standards enabling them to translate their findings into dollar amounts....Based on this research we can expect relatively uniform assessments of compensatory damages in tobacco cases, but widely variant damages that will be appreciably higher when awarded by local juries than by juries in a national class action...⁹

In *TVT Records v. Island Jam Music* one of the articles that forms a part of *Punitive Damages* is quoted as authority.¹⁰

Critiques of individual articles in the Exxon body of research have appeared in the academic literature.¹¹ An amicus brief submitted in *State Farm v. Campbell*¹² signed by twenty-one social scientists and legal scholars, including the present author, one of the drafters of the brief, offered their professional opinion that the broad policy claims of the Exxon-funded authors were not justified.¹³

Yet, for several reasons there is a serious danger that these critiques will be ignored by judges, legislators and other policy makers. First, the “outlier” awards involving high punitive to

⁷ Brief of Certain Leading Business Corporations as Amici Curiae in Support of Petitioner, *State Farm Mutual Automobile Insurance Company v. Campbell*, 538 U.S. 408 (2003).

⁸ *Cooper Industries, Inc v. Leatherman Tool Group, Inc.*, 532 U.S. 424, 439 (2001)

⁹ *In re Simon II Litigation*, 211 F.R.D. 86,106 (2002). Punitive Damages if further cited at 110 and 161

¹⁰ *TVT Records v. Island Jam Music*, 279 F. Supp. 2d 413 (2003): at 417 quoting Sunstein, Kahneman, and Schkade, *Assessing Punitive Damages (With Notes on Cognition and Valuation)* 107 YALE LAW JOURNAL 2071 (1998) “...juries assess punitive damages in wholly unpredictable amounts bearing no necessary relation to the actual harm caused;” and at 430: “[A]risk of extremely high awards is likely to produce excessive caution in risk-averse companies. Hence unpredictable awards create both unfairness and ...inefficiency, in a way that may overdeter desirable activity.”

¹¹ see Richard Lempert, *Juries, Hindsight Bias, and Punitive Damage Awards: Failures of a Social Science Case for Change*, 48 DEPAUL LAW REVIEW 867, 877 (1999); Neil Vidmar, *Juries Don’t Make Legal Decisions! And Other Problems: A Critique of Hastie et al. on Punitive Damages*, 23 LAW AND HUMAN BEHAVIOR 705 (1999); Robert MacCoun, *Epistemological Dilemmas in the Assessment of Legal Decision Making*, 23 LAW AND HUMAN BEHAVIOR 723 (1999); Steven Garber, *Punitive Damages and Deterrence of Efficiency-Promoting Analysis: A Problem Without A Solution* 52 STANFORD LAW REVIEW 1809 (2000); Robert MacCoun, *The Costs and Benefits of Letting Juries Punish Corporations: Comment on Viscusi*, 52 STANFORD LAW REVIEW 1821,1827 (2000).

¹² 123 S.Ct. 1513 (2003).

¹³ Brief of Amici Curiae of Certain Leading Social Scientists and Legal Scholars in Support of Respondents, *State Farm* (No.01-1289). Professor Theodore Eisenberg and I drafted the amicus. Many of the ideas and critiques in this article overlap with parts of that amicus. However, the views that I express in this article are solely my own; I make no representation that the co-signers of the amicus necessarily agree with my views expressed in this article.

compensatory damages ratios that have been caused the Supreme Court to struggle with legal doctrine and that have garnered so much attention in the mass media may be seen as the result of typical jury behavior that is consistent with the claims of *Punitive Damages* and its authors. Second, the eminence of the authors and their prolific outpouring of articles may create the impression of general scientific consensus about their conclusions. Third, the authors' claim that the studies involved "thousands of jury-eligible citizens, on hundreds of mock juries, and on a few hundred experienced trial judges asked to render punitive damages and verdicts under realistic task conditions"¹⁴ is likely to preempt close scrutiny of the research methodology.¹⁵ Similar to a failure to read the fine print in a contract, judges and legislators may say "grasshopper" and hastily jump ahead to the conclusions of *Punitive Damages*.¹⁶ They will be in error.

Two important reviews of *Punitive Damages* have recently appeared. Professor Catherine Sharkey has challenged whether the authors of the book have sufficiently considered the societal deterrence and compensatory functions of punitive damages.¹⁷ These goals of punitive damages have been minimized, ignored or rejected in the simulation experiments. She also pointed out that the authors neglect the role of institutional context and appellate review in controlling extreme jury awards.¹⁸ Based on her analysis she argued that, while the authors of *Punitive Damages* "have done an admirable job of producing and analyzing empirical data, their policy prescriptions go too far in using that data to castigate the jury, and not far enough in suggesting logical reforms within our existing system."¹⁹ She also argued that the "call to banish the jury in the realm of punitive damages is at best premature, at least on the basis of their empirical findings."²⁰

¹⁴ PUNITIVE DAMAGES at 211.

¹⁵ The sheer number of experiments as well as the number of subjects used in the research are likely to impress even careful scholars and judges. For instance, Professor Catherine Sharkey's review of *Punitive Damages*, while critical of aspects of the research, states that "[t]he sheer scope the collective research endeavor is unprecedented," and "draws upon experimental research based upon twenty experimental studies involving more than 8,000 jury-eligible citizens and more than 600 mock juries,." Catherine Sharkey, *Punitive Damages: Should Jurors Decide?* 82 TEXAS LAW REVIEW 382, 386 (2003).

¹⁶ This is not to say that judges are incapable of reading the methodology or putting it into context, merely that their attention is centered on the legal issues. However, as I will show, some of the methodological issues are subtle and might easily escape even a careful judge not trained in social science research.

¹⁷ Catherine Sharkey, *Punitive Damages: Should Jurors Decide?* 82 TEXAS LAW REVIEW 382,400 (2003); see also Catherine Sharkey, *Punitive Damages as Societal Damages*, 113 YALE L. J. 347 (2003).

¹⁸ Sharkey, *Punitive Damages* at 404. See also, Neil Vidmar, *supra* note 11 at 711. In his introduction to PUNITIVE DAMAGES, Professor George Priest briefly acknowledges judicial review in a footnote, see PUNITIVE DAMAGES at 16, note 17.

¹⁹ Sharkey, *Should Jurors Decide*, *supra* note 17 at 383. (Scientists use the word data as a plural and datum to refer to the singular but Sharkey's reference to "that data" rather than "those data" reflects a contemporary trend in lay usage of "data".)

²⁰ *Id.* at 385.

Professor Neal Feigenson conducted an extensive review of *Punitive Damages*.²¹ He noted that the book should have given greater attention to real world research on punitive damages and to the normative criteria against which jury decisions should be judged. He also documented shortcomings of some of the experiments. While acknowledging that many of the findings from the experiments seem consistent with other research findings, he concluded that “...in various ways, generally small in themselves but cumulatively hard to ignore –sometimes in the design of the experiments, more frequently in the language employed to characterize their findings and in the selective use (or more often, neglect) of other empirical data to provide a context for their findings—the authors’ normative position leads them to present their results tendentiously.”²²

Feigenson’s critique also offered the view that *Punitive Damages*’ emphasis on optimal deterrence to the exclusion of retribution as an additional purpose of punitive damages limits the usefulness of its conclusions.²³

The Sharkey and Feigenson articles should be read carefully because they present a compelling critique of aspects of the research in *Punitive Damages*. I am in agreement with almost all of their basic conclusions, but in this article I go much further.

First, I draw attention to Feigenson’s observation that there are data from the real world that do not support the broad conclusion that juries are incompetent and unreliable and anti-business.²⁴ The failure of the authors of *Punitive Damages* to address this body of data is a serious shortcoming. I then turn to critical aspects of the simulation experiments that are not sufficiently engaged in the Feigenson and Sharkey reviews, namely the extent to which the simulation experiments in *Punitive Damages* replicate the real world of the jury trial and the context in which juries make decisions, a problem that social scientists label “ecological validity.” As my analysis will show, the evidence and arguments in the simulation “trials” were not only very brief but,

²¹ Neal Feigenson, Can Tort Juries Punish Competently? 78 CHICAGO-KENT LAW REVIEW 101 (2002).

²² Id. at 105.

²³ Id. at 147.

²⁴ Id. at 244, 250-5, 255-56. Throughout the program of research various of the authors have asserted or implied that jurors are anti-business. See PUNITIVE DAMAGES at 64, 73, 113-14, 233. In an article published prior to the book Hastie et al, A Study of Juror and Jury Judgments in Civil Cases: Deciding Punitive Damages, 22 LAW AND HUMAN BEHAVIOR 287 (1998), a conclusion Vidmar, supra note 11 at 712, I found not supportable from their data. I do not address this matter further in this article but the reader is referred generally to Valerie Hans, BUSINESS ON TRIAL (2000), especially Chapters 5, 6 and 7, for research showing that jurors are not inherently anti-business nor do they make compensatory awards based on the wealth of the defendant, the so-called “deep pockets” effect though they do sometimes hold businesses to higher standards of accountability. See also, Robert MacCoun, Differential Treatment of Corporate Defendants by Juries: An Examination of the “Deep Pockets” Hypothesis, 30 LAW AND SOCIETY REVIEW 121 (1966); Neil Vidmar, MEDICAL MALPRACTICE AND THE AMERICAN JURY: CONFRONTING THE MYTHS ABOUT JURY INCOMPETENCE, DEEP POCKETS, AND OUTRAGEOUS DAMAGE AWARDS (1995) at Chapter 18, pp.203-220.

additionally, excluded features of real trials, such as arguments and counter-damage proposals by the defendant. Further, there are serious problems with the design of some of the experiments. The imbalance of information provided to the simulating jurors overwhelmingly favored the plaintiff and undermined the various experiments' ability to allow scientifically appropriate causal inferences. These omissions and commissions raise serious doubts about conclusions that can be drawn from *Punitive Damages*.

Before turning to the task of examining the methodology in *Punitive Damages* two matters need to be emphasized as strongly as possible. First, this article does not take issue with the Supreme Court's attempts to address the issue of "outlier" punitive damages awards, particularly the concern with the ratios between punitive and compensatory damages.²⁵ I leave doctrine to judges, except insofar as the doctrine is based on false empirical assumptions.

Second, this article is not a condemnation of experimental simulation research.²⁶ Simulation experiments can shed important light on basic individual or group decision processes.. Sometimes they are the only practical way to investigate an important legal issue; combined with other data, they can yield data having direct policy or fact implications.²⁷ The issue regarding *Punitive Damages* is whether it is permissible to make broad sweeping policy statements from the simulation research reported in the volume, no matter how many experiments were undertaken or how prominent the authors are, when those data are contradicted by real world data, when the experiments leave out or distort features that are inherent parts of real world juries' decision-making contexts and when there are confounds in the experimental designs that prejudice the researchers' ability to draw valid causal inferences.

A comment on terminology. Throughout this article many references are made to the persons who participated as subjects in research studies. There are many conventions for labeling

²⁵ See *BMW of N. Am., Inc. v. Gore*, 517 U.S. 559 (1996); *Cooper Tool Industries Inc v. Leatherman Tool Group, Inc.* 121 S. Ct. 1678 (2001); *State Farm v. Campbell* ; See also Marcia Coyle, New Battles Come over Punitives: High Court's Guides Include Ratios 25 NATIONAL LAW JOURNAL PA1, April 14, 2003.

²⁶ Indeed, my own vita contains a substantial number of published studies involving jury simulations and I have cited experimental simulation research in legal testimony when it is consistent with other evidence: see , e.g. Neil Vidmar, MEDICAL MALPRACTICE AND THE AMERICAN JURY: CONFRONTING THE MYTHS ABOUT JURY INCOMPETENCE, DEEP POCKETS, AND OUTRAGEOUS DAMAGE AWARDS, U. of Michigan Press , 1995 at Chapter s 18 and 19 and Vidmar Case Studies of Pre- and Midtrial Prejudice in Criminal and Civil Litigation, LAW AND HUMAN BEHAVIOR, 26, 73 (2002) <http://www.law.duke.edu/pub/vidmar/pretrialPublicity.pdf>

²⁷ E.g. see, Shari Diamond and Judith Levi, Improving Decisions on Death by Revising and Testing Jury Instructions. 79 JUDICATURE 224 (1996); Gary L. Wells et al., Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads, 22 LAW & HUMAN BEHAVIOR 603 (1998).

these persons: “subjects,” “respondents,” “simulating jurors,” “jurors” and “persons.” I use these various terms interchangeably.

II. Punitive Damages in the Real World

One of the most striking observations that can be made about *Punitive Damages* is the paucity of references the authors make to real world studies of the incidence and magnitude of punitive damages. More than a few such studies do exist and one cannot but wonder why those findings are ignored, particularly since they do not lend support to *Punitive Damages*’ broad claims about a widespread problem of frequent and “incoherent” judgments. These real world studies have shown that punitive awards are generally infrequent. When punitive awards are rendered, they most often occur in intentional tort cases involving financial matters. Generally, punitive awards are modest in comparison to compensatory awards.

One of the reasons that these studies may not have received more attention from judges, legal scholars and commentators is that they are usually presented in the form of rather austere aggregate statistics, often accompanied by regression analyses or other arcane, albeit important, methods of assessing data sets.

Mary Rose and I described data from Florida state courts between 1989 and 1998 that attempted to put a more concrete face on real world punitive damage awards by breaking the data into categories, comparing ratios of punitive damages to compensatory damages and describing the nature of “outlier awards.”²⁸ Our data showed that the actual number of punitive damage awards in Florida from 1989 through 1998 averaged about 23 cases per year.²⁹ Moreover, slightly over 23 percent involved modest punitive awards against impaired drivers.³⁰ The next biggest category involved claims of fraud and other financial losses (just over 17 percent) and the third largest number of cases involved claims of physical and sexual assaults (about 16 percent).³¹ Product liability cases accounted for only slightly over 7 percent of cases and a closer examination of these cases revealed that most involved asbestos claims from the early part of the 1990s decade.³² Responding to the legal controversy involving the ratio between punitive and compensatory

²⁸ Vidmar and Rose, Punitive Damages by Juries In Florida: *In Terrorem* and In Reality, 38 HARVARD JOURNAL ON LEGISLATION, 487 (2001).

²⁹ Id. at 492.

³⁰ Id. at 495.

³¹ Id.

³² Id. at 496.

awards,³³ we examined the data further to assess the ratio of punitive awards to compensatory awards by types of cases.³⁴ These data are reproduced in Table 1.³⁵

³³ See note 26 *supra*.

³⁴ Vidmar and Rose, *supra* note 29 at 501. The ratios reported in this column are the medians of the individual ratios.

³⁵ *Id.* at 501, designated as Table 3 in the original article. Research currently in progress will expand this data set and further examine the relationship between punitive and compensatory damages in greater detail.

Table 1. Distribution of median total and punitive awards and the median ratio of punitive to compensatory awards, by case type (1988 – 2000)

Type of case	N cases	Median Total Award	Median Punitive Award	Pun: Comp Ratio (median)
Motor vehicle accidents	63	\$284,736	\$21,579	0.1:1
Fraud, contract violation, and other financial damage cases	47	\$392,158	\$318,055	1.0:1
Assaults	43	\$221,461	\$59,832	0.4:1
Products liability	20	\$2,245,635	\$666,936	0.8:1
Information violations (privacy, slander, defamation, libel)	20	\$191,264	\$108,530	1.1:1
False imprisonment/false arrest	20	\$234,752	\$139,814	0.4:1
Premises liability	17	\$933,660	\$200,081	0.5:1
Discrimination/ harassment	13	\$1,344,841	\$1,030,530	2.3:1
Professional negligence	12	\$3,078,133	\$1,006,172	2.5:1
Workplace injuries/ failure to pay benefits	11	\$317,260	\$71,820	0.5:1
Other: Improper treatment of dead persons	4	\$3,434,572	\$3,052,075	6.3:1
Overall	270	\$612,028	\$151,871	0.7:1

Note: Awards adjusted to 1999 dollars.

Table 1 shows that the median ratios ranged from 0.1:1 in impaired driver accidents to 6.3:1 in four cases involving improper treatment of deceased persons. The median punitive damages ratio in products liability cases was 0.8:1. Considering all cases, the average punitive damages award was only 68 percent of the compensatory award.³⁶

Examination of the claims made by the plaintiffs in most of the cases led to the conclusion that there was a prima facie case for punitive damages.³⁷ For instance, in one case the plaintiff alleged that a company, cited previous time for dumping toxic chemicals, dumped chemicals in a dumpster that killed two young boys.³⁸ In another case the plaintiff alleged that the captain of a cargo ship ignored warnings about dangerous levels of carbon monoxide and failed to inform the Coastguard about the condition, killing one of the rescuers who had come to the aid of a crewman.³⁹ Other cases were similar. Businesses engaging in commonly accepted practices or those taking normal safety precautions were rarely assessed punitive damages.⁴⁰ Moreover, when both a business and an individual were jointly named in a lawsuit, the business was often found liable for compensatory damages but in 36 percent of these cases only the individual defendant was assessed punitive damages, suggesting that the juries differentiated between behaviors of defendants and were not necessarily anti-business.

The use of the median statistic can obscure the presence of large awards, and indeed there were some very large awards in some cases.⁴¹ Yet, a closer examination of the top twenty awards yielded interesting insights. Six of these “mega” awards involved cases in which the defendants were not represented by a lawyer and a seventh case involved a company already in bankruptcy and facing criminal charges. Other awards were denied in post-trial review or subjected to *remittitur*, and at the time of the reports, still other cases had pending motions about the award.⁴² Overall, at least half of the mega awards resulted in no payment to the plaintiff or a reduced payment.⁴³

³⁶ Id. at 503.

³⁷ Id. at 496-499.

³⁸ *Perez v. William Recht Co.*, No.92-8983, 1995 WL 861061 (Fla. Cir. Ct. Sept 28, 1995)

³⁹ *Tabil v. Juno Marine Agency, Inc.*, No. 88-45327-CA-22, 1998 WL 355212 (Fla Cir.Ct., Feb 23, 1998).

⁴⁰ Vidmar and Rose, *supra* note 29 at 499.

⁴¹ In one instance the jury awarded \$1 in compensatory damages and \$14,000 in punitive damages, a 14,000:1 ratio of punitive to compensatory damages. As this example illustrates one or two aberrant cases can distort statistical means and standard deviations. There are statistical ways to control for such extreme values and these will be explored in future research with the data.

⁴² Id. at 503-506.

⁴³ Id. at 506.

Other Real World Research Studies

While our Florida data present a detailed picture of only one state jurisdiction during a span of a dozen years, a substantial number of additional studies collected by various independent researchers in private and government institutions and in universities⁴⁴ have produced findings generally consistent with the Florida data reported above.⁴⁵ While it is not my purpose here to review this research in detail, a few examples are illustrative. Michael Rustad studied twenty-five years of punitive damages in products liability cases.⁴⁶ His findings indicated that punitive verdicts are rare, tend to be proportionate to the wrongdoing and tend to be based upon avoidable patterns of corporate recklessness.⁴⁷ A U.S. Bureau of Justice Statistics study indicated that in 1996 the median punitive damage award for all jury trial cases nationwide was \$50,000.⁴⁸

The RAND Institute for Civil Justice studied punitive damages occurring in selected jurisdictions for a period extending from 1960 through 1994.⁴⁹ Researchers in that organization concluded that punitive damages were awarded rarely, and mostly in business disputes and

⁴⁴ Rustad, In Defense of Punitive Damages in Products Liability: Testing Tort Anecdotes with Empirical Data, 78 IOWA L. REV. 1 (1992); Rustad and Koenig, Reconceptualizing Punitive Damages in Medical Malpractice: Targeting Amoral Corporations, Not "Moral Monsters," 47 RUTGERS L. REV. 975 (1995); Peterson, et al. Punitive Damages: Empirical Findings, Institute for Civil Justice. RAND Corporation (1987); Rustad and Koenig, Reconceptualizing Punitive Damages in Medical Malpractice: Targeting Amoral Corporations, Not "Moral Monsters," 47 RUTGERS L. REV. 975 (1995); Rottman et al., A Step Above Anecdote: A Profile of the Civil Jury in the 1990's, 79 JUDICATURE 223 (1996); C. DeFrances and M. Litras Civil Trial Cases and Verdicts in Large Counties, 1996, Bureau of Justice Statistics Bulletin, September, 1999; Eisenberg *et al.*, *The Predictability of Punitive Damages*, 26 J. OF LEGAL STUDIES 623 (1997); Moller, Trends in Civil Jury Verdicts Since 1985, The Institute for Civil Justice, RAND Corporation (1996); Daniels and Martin, CIVIL JURIES AND THE POLITICS OF REFORM (1995); Moller, Pace and Carroll, Punitive Damages In Financial Injury Jury Verdicts, 28 JOURNAL OF LEGAL STUDIES 283 (1999); Eisenberg, La Fountain, Rottman, and Wells, Juries, Judges, and Punitive Damages: An Empirical Study, 87 CORNELL LAW REVIEW 743 (2002); Koenig, The Shadow Effect of Punitive Damages on Settlements, 1998 WISC. L. REV. 169 (1998);⁴⁴ Karpoff and Lott, On the Determinants and Importance of Punitive Damage Awards, XLII JOURNAL OF LAW AND ECONOMICS 527 (1999); Rustad, Unravelling Punitive Damages: Current Data and Further Inquiry, 1998 WISCONSIN LAW REVIEW 15 (1998).

⁴⁵ One of these studies, by Karpoff and Lott, *supra* note 44, assessed the magnitude of punitive damages against the type of lawsuit and the defendant company, variables accounting for only between one and two percent of the variance (at 571), causing the authors to conclude that punitive awards are highly variable and unpredictable. However, there were only two variables and they were such gross measures that this conclusion cannot be taken seriously. On the other hand it is important also to note that these same authors compared punitive awards to compensatory awards; the punitive award was positively and significantly related to the compensatory award (at 543).

⁴⁶ Michael Rustad, *supra* note 44..

⁴⁷ For a review and summary of nine studies actual jury verdicts in punitive damages cases see, Thomas Koenig and Michael Rustad, IN DEFENSE OF TORT LAW (2001) at 180-184..

⁴⁸ C. DeFrances and M. Litras CIVIL TRIAL CASES AND VERDICTS IN LARGE COUNTIES, 1996, Bureau of Justice Statistics Bulletin, September, 1999.

⁴⁹ Peterson, et al. PUNITIVE DAMAGES: EMPIRICAL FINDINGS, Institute for Civil Justice. RAND Corporation (1987). Moller, TRENDS IN CIVIL JURY VERDICTS SINCE 1985, The Institute for Civil Justice, RAND Corporation (1996).

intentional tort cases, and that, generally, the awards were modest in relation to compensatory damages. Another Institute For Civil Justice report covered the period 1985 to 1994 and involved 15 state courts of general jurisdiction.⁵⁰ With respect to punitive damages, the report concluded: “[p]erhaps the most striking finding that emerges from the jury verdict data in this study is that punitive damages are awarded very rarely.” The report also concluded: “The discussion about punitive damages focuses primarily on products liability, but in jurisdictions we examined, most punitive damages were awarded in intentional tort and business cases....” and “[i]n contrast, products liability was the underlying cause of action in only 4.4 percent of the punitive damage awards made.”

Stephen Daniels and Joanne Martin at the American Bar Foundation examined punitive damages from a number of locations around the United States over twenty-one years. While finding some variations across jurisdictions they concluded that juries did not routinely award punitive damages, and that median awards tended to be modest.⁵¹

In 1998 Michael Rustad reviewed nine studies based on real world empirical data.⁵² He observed that there are many things that are unknown about punitive damages but concluded that

Every empirical study of punitive damages demonstrates that there is no nationwide punitive damages crisis. The research shows that punitive damages cluster in business tort and intentional tort cases, not personal injury. The increase in punitive damages is largely confined to a few jurisdictions.⁵³

Eisenberg et al. compared judge versus jury trial outcomes in 45 of the nation's largest counties in 1996.⁷⁷ After controlling for differences in the types of cases heard by each set of decision makers, those authors concluded that there was no substantial evidence that judges and juries differed neither in the rate at which they awarded punitive damages nor in the basic ratio of punitive to compensatory damages. Jury trials did have a greater range of punitive damages for a given level of compensatory damages but in the end, those authors concluded, there were only a “trivially” few cases in which the jury award would have exceeded what a judge might have awarded.

⁵⁰ Moller, TRENDS IN CIVIL JURY VERDICTS SINCE 1985, The Institute for Civil Justice, RAND Corporation (1996).

⁵¹ Daniels and Martin, *supra* note 44 at 238-243.

⁵² Rustad, Unravelling Punitive Damages: Current Data and Further Inquiry, 1998 WISCONSIN LAW REVIEW 15 (1998).

⁵³ *Id.* at 69.

⁷⁷ Eisenberg, et al., *supra* note 44.

Some of the real world studies have also addressed the issue of post-verdict adjustments of punitive awards.⁵⁴ Generally, these studies found that very large verdicts were adjusted downward by trial or appellate judges or through settlements. Indeed, it is worthwhile noting that these findings are consistent with a study by one of the authors of *Punitive Damages*. Professor Viscusi concluded that in products liability cases involving punitive damages

“plaintiffs received only 29% of the original punitive award. Courts often reduce punitive damages on appeal, and defendants may negotiate a reduction in this amount in return for prompt payment of the damages amount.”⁵⁵

Contradicted Claims

The real world studies and the conclusions drawn by their respective authors have been widely available to researchers. Yet, while hardly acknowledging these adverse data, *Punitive Damages* makes the grand claim in Chapter 1:

To our knowledge, there *is not a single instance* in which our results disagree with findings from other experiments conducted by independent groups of behavioral researchers *or with any findings from the statistical analysis of actual trial verdicts*.⁵⁶ (italics added)

In Chapter 12 summarizing the series of studies the claim is modified slightly:

Support for the claim that dollar awards are erratic has accumulated in research based on analyses of archival award statistics demonstrating the unpredictability of jury punitive damages awards in actual courtrooms. Our reading of this controversial literature is that jury verdicts are highly variable for similar cases, especially across jurisdictions.⁵⁷

One study cited in Chapter 12 and discussed again in Chapter 13,⁵⁸ was to an empirical study by Eisenberg et al.⁵⁹ The abstract to that article reads, in part:

⁵⁴ Peterson, *et al.* PUNITIVE DAMAGES: EMPIRICAL FINDINGS, Institute for Civil Justice, RAND Corporation (1987); Rustad, In Defense of Punitive Damages in Products Liability: Testing Tort Anecdotes with Empirical Data, 78 IOWA L. REV. 1 (1992); Rustad and Koenig, Reconceptualizing Punitive Damages in Medical Malpractice: Targeting Amoral Corporations, Not “Moral Monsters,” 47 RUTGERS L.REV. 975 (1995); See generally, Thomas Koenig, The Shadow Effect of Punitive Damages on Settlements 1998 WISCONSIN LAW REVIEW 157, 202-207 (1998).

⁵⁵ W. Kip Viscusi, REFORMING PRODUCTS LIABILITY (1991) at 94.

⁵⁶ PUNITIVE DAMAGES at 20.

⁵⁷ PUNITIVE DAMAGES at 213-214.

⁵⁸ PUNITIVE DAMAGES at 245-248. Chapter 13 also refers to an additional study: Jonathan Karpoff and John Lott, On the Determinants and Importance of Punitive Damage Awards, 42 JOURNAL OF LAW AND ECONOMICS, 527 (1999). That research attempted to explain the variation in punitive awards with two variables, the type of case and the size of the defendant company, finding little relationship with the size of the award. The research did not examine specific case facts such as the egregiousness of the defendant’s conduct. It also attempted to compare cases in which punitive

...this article shows a strong and statistically significant correlation between compensatory and punitive damages. ... In addition we find no evidence that punitive damages awards are more likely when individuals sue businesses than when individuals sue individuals. With respect to award frequency, juries rarely award punitive damages and appear to be especially reluctant to do so in the areas of law that have captured the most attention, products liability and medical malpractice. Punitive damages are most frequently awarded in business/contract cases and intentional tort cases. The frequency-of-award findings are consistent with all major studies of punitive damages.⁶⁰

While the authors of *Punitive Damages* are entitled to interpret studies differently than the original authors of these various real world studies, it is peculiar that they did not acknowledge the studies and give reasons why the findings should be interpreted differently.⁶¹

III. Methodological Issues in Simulation Research: Context for Interpreting the Experiments in *Punitive Damages*

Real world data derived from archival analyses do not provide direct insights into jury decision-processes. They provide only the end results. Comparing the results with other variables allows us to make inferences about how jurors behave, but the inferences are indirect. One way to discover how juries decide punitive damages would be observations of juries as they deliberate, but with one notable exception this has not been allowed for legal policy reasons.⁶² Another way is to randomly assign some real world juries to deliberate under some instructions or conditions

“awards” were reflected in settlements, a dubious undertaking since the characteristics of settled cases are so different from cases that go to trial, see generally, Michael Saks, Do We Really Know Anything About the Behavior of the Tort litigation System –and Why Not?, 140 UNIVERSITY OF PENNSYLVANIA LAW REVIEW 1147 (1992); Vidmar N. Pap and Circumstance: What Jury Verdict Statistics Can Tell Us about Jury Behavior and the Tort System. 27 SUFFOLK UNIVERSITY LAW REVIEW 1205 (1994/1996).

⁵⁹ Eisenberg et al., supra note 44.

⁶⁰ Id. at 623.

⁶¹ Feigenson, supra note 21 at 250 and 255-6 also raises similar points about the failure of the authors of PUNITIVE DAMAGES to confront data bearing on actual jury verdict on punitive damages.

⁶² See Diamond, Vidmar, Rose, Ellis and Murphy, Juror Discussions During Trials: Studying an Arizona Innovation, 45 ARIZONA LAW REVIEW 1 (2003); Diamond and Vidmar, Jury Room Ruminations on Forbidden Evidence, 87 VIRGINIA LAW REVIEW 1857 (2001). Systematic post-verdict interviews with jurors is still another method, but it has problems involving the accuracy and reliability of juror recall, see Paula L. Hannaford et al., Permitting Jury Discussions During Trial: Impact of the Arizona Reform, 24 LAW & HUM. BEHAV. 359 (2000); Paula L. Hannaford-Agor et al., “Speaking Rights”: Evaluating Juror Discussions During Civil Trials, 85 JUDICATURE 237 (2002).

while others deliberate under different conditions, but this too, usually conflicts with legal policy.⁶³ Even in the extremely rare instances in which juries deciding real cases have been studied, problems of differences of confounding variables within trials and between different trials remain. Although these confounds can be minimized through various statistical measures, an alternative way of gaining knowledge is through simulation research.

Experimental simulations involving human behavior, including jury behavior, are widely used in the behavioral sciences and in other disciplines as well. *Punitive Damages* is correct in its assertion that systematic experimentation is a “venerable and widely used method of scientific investigation,”⁶⁴ and “used routinely in engineering, medical and behavioral research.”⁶⁵ Experimental methods allow causal inferences by controlling variables that are confounded in the real world.⁶⁶ However, applying data from simulation studies to real world problems, particularly problems involving broad policy changes, presents complications that require a great deal of caution, including close attention to methodological details. Consider three important social science concepts: internal validity, external validity and ecological validity.

Social scientists refer the extent to which extraneous variables are controlled as the problem of *internal validity*.⁶⁷ The more extraneous variables are controlled the more likely that a relationship between one variable and another can be inferred to be causal. Experimental studies of jury simulations can control variables and thus have the potential to be high in internal validity. For example, if we are concerned about whether one form of jury instruction is superior to others different instructions can be given to different sets of simulating jurors who hear the same summary of trial evidence; then the jurors’ comprehension under the differing instructions can be assessed. Because all of the trial evidence and the form and manner in which it is presented is exactly the same across trials, any differences between types of instructions can be assumed to be due to the instructions. Alternative explanations can be ruled out. A causal link is established and

⁶³ In the two Arizona jury studies, Diamond et al., supra note 62 and Hannaford et al. supra note 62 juries were randomly assigned to different sets of instruction conditions, but such experiments have been rare in jury research.

⁶⁴ PUNITIVE DAMAGES at 18.

⁶⁵ Id. at 19.

⁶⁶ See William R. Shadish, Thomas D. Cook and Donald T. Campbell, EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE (2002) at 33-102; Elliot Aronson, Timothy Wilson and Marilynn Brewer, Experimentation in Social Psychology, in Daniel Gilbert, Susan Fiske and Gardner Lindzey, eds., THE HANDBOOK OF SOCIAL PSYCHOLOGY, Vol I (1998), at 129-135; Marilyn Brewer, Research Design and Issues of Validity, in Harry T. Reiss and Charles M. Judd, eds., HANDBOOK OF RESEARCH METHODS IN SOCIAL AND PERSONALITY PSYCHOLOGY (2000) at Chapter 1, pages 3-39.

⁶⁷ Brewer, supra note 66 at 4; Shadish et al., supra note 66 at 53-63.

we can say the experiment is high in internal validity.⁶⁸ But achieving high internal validity usually has costs. By controlling some variables the interplay of other variables as they occur in the real world and interact with one another is often altered.

Thus, even if a causal link is established in a particular experiment, we can ask if the results are robust enough that they transcend the subject population or the particular trial evidence or the particular moment in time. If the above experiment on jury instructions used college students as jurors would we get the same results with a non-student group of jury-eligible voters? Would the results be the same using jury-eligible persons from Texas as opposed to jury-eligible persons from Massachusetts? Would they obtain if the case was changed, say, from a medical malpractice case to a case involving investment fraud? Would an experiment involving a criminal defendant identified as being from an Arab country yield different results if it was carried out prior to September 11, 2001 as opposed to after that date?⁶⁹ To the extent that the results generalize across these various settings, the study may be seen to have *external validity*.⁷⁰

Concerns about external validity can be partially overcome when the simulation experiments are conducted in a variety of different settings or under different conditions showing a certain amount of consistency, or generalizability. This, in fact, is one of the claims made by the authors of *Punitive Damages*, namely that they have achieved “repeated replication of the research results across experimental tests within the research program, and from replication in independent research programs.”⁷¹

But there is a third validity issue that is of concern. All simulation experiments, whether in the social, biological or physical sciences, are only analogues. They differ in the degree to which there is verisimilitude between the experimental context and the real world. Some jury experiments, for example, have presented respondents with only a short one- or two-page description of trial evidence in order to assess the effects of a certain variable that was hypothesized to have an influence on jury decision-making. One such experiment presented groups of jurors and groups of judges with brief trial concerning a products liability case. It varied conditions bearing on inadmissible evidence in order to test whether judges were as susceptible to inadmissible evidence

⁶⁸ Often the question is posed as to whether a plausible alternative hypothesis can explain the relationship between the variables. If not the study is said to be high in internal validity, see William Shadish, Thomas Cook and Donald T. Campbell, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE*, 53-64 (2002).

⁶⁹ For further discussion on this example, see Vidmar, *When All of Us Are Victims: Juror Prejudice and “Terrorist” Trials*, 78 *CHICAGO-KENT LAW REVIEW* 1143 (2003).

⁷⁰ Shadish et al., *supra* note 66 at 83-102. Brewer, *supra* note 66 at 10-12.

⁷¹ *PUNITIVE DAMAGES* at 20.

as jurors.⁷² (They were !) Other experiments attempt to mimic the real world as closely as possible within certain constraints. Thus, one experiment exploring the effect of trial procedures on punitive damages created a trial script based on an actual trial that contained all the elements of the trial, including presentations by the plaintiff and defendant and instructions by the judge.⁷³ Different versions of parts of the trial were created to test hypotheses about weak or moderately strong evidence favoring the plaintiff and about the effects of unitary or bifurcated trials.⁷⁴ Professional actors played the roles of the lawyers and the judge in a courtroom and a two and one-half hour videotape trial was produced. Jury-eligible persons were recruited to serve as jurors. These simulating jurors were exposed to a form of voir dire, watched the trial and deliberated to a verdict. Although differences in the degree of verisimilitude was quite substantial, both experiments described above produced useful knowledge.

Nevertheless, even the most complex simulation does not create all the conditions of the real world. Applying its results directly to policy must be treated with some caution, even if the findings are replicated in other experiments. Most social science researchers explicitly recognize this issue in their articles, the problem of *ecological* validity.

As described by Professor Marilynn Brewer in a leading social psychology research manual:

The question of whether an effect holds up across a wide variety of people or settings is somewhat different than asking whether the effect is representative of what happens in everyday life. This is the essence of ecological validity—whether an effect has been demonstrated to occur under conditions that are typical for the population at large.

⁷² Stephan Landsman and Richard Rakos, A Preliminary Inquiry into the Effect of Potentially Biasing Information on Judges and Jurors in Civil Litigation, 12 BEHAVIORAL SCIENCES AND THE LAW 113 (1994).

⁷³ Stephen Landsman et al., Be Careful What You Wish For: The paradoxical Effects of Bifurcating Claims for Punitive Damages 1998 WISCONSIN LAW REVIEW 297, 309-314; Shari Diamond, et al., Juror Judgments About Liability and Damages: Sources of Variability and Ways to Increase Consistency, 48 DEPAUL LAW REVIEW 301,303-305 (1998). For other complex simulation experiment attempting to mimic real world trial conditions see: Irwin Horowitz and Kenneth Bordens, The Effects of Outlier Presence, Plaintiff Population Size, and Aggregation of Plaintiffs on Simulated Civil Jury Decisions, 12 Law and Human Behavior 209 (1988) and Horowitz et al., Effects of Trial Complexity on Decision Making, 81 JOURNAL OF APPLIED PSYCHOLOGY 757 (1992).

⁷⁴ Despite its apparent relevance to experimental studies of punitive damages, the Landsman et al. study is not referenced in PUNITIVE DAMAGES. Some of the findings of that study might be interpreted as supporting the Exxon research, but many findings do not. For instance, Dimond et al., supra note 73 at 316, found that that the standard deviations of damages awards by the deliberating juries to be much smaller than those of pre-deliberation awards of the individual jurors. In comparing any of the experiments reported in PUNITIVE DAMAGES to the Landsman study with regard to an attempt to create real world trial conditions PUNITIVE DAMAGES does not fare well.

Representativeness is not the same as robustness. Generalizability in the robustness sense asks whether an effect can occur across different settings and people; ecological validity asks whether it [a finding or set of findings] does occur in the world as it is.⁷⁵

It is for this reason that writings about the theory of social psychology emphasize that experimentally derived data, regardless of how many experiments are conducted, must be checked against the real world, especially when policy recommendations are made from the research.⁷⁶ Thus, to quote Brewer again:

When the research is essentially descriptive, ecological validity may be essential. When the purpose is utilitarian, robustness of an effect is particularly critical. The fragility and non-generalizability of a finding may be a fatal flaw if one's goal is to designing an intervention to solve some applied problem.⁷⁷

Lawyers will quickly see that ecological validity has a close similarity to the relevancy requirement of F.R.E. 702, as articulated in the *Daubert* cases.⁷⁸ Particularly germane is the second prong of Rule 702, namely the degree of “fit” between scientific evidence and the contested issues in the case.⁷⁹ No matter how rigorous the methodology and number of replications (producing internal validity and external validity) the research must have “fit” to the problem at hand. In the remand decision, commonly called *Daubert II*, evidence derived from structure activity, in vitro and in vivo experiments was judged to not have the same “fit” in proving

⁷⁵ Brewer, supra note 66 at 12. Some research authorities prefer to treat ecological validity, as Brewer and I use it here, as an aspect of external validity, see William Shadish, Thomas Cook and Donald Campbell, EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE, 37-39(2002) but the ecological validity is a better concept for conveying application of research in policy contexts.

⁷⁶ Indeed, one of the authors of PUNITIVE DAMAGES has previously acknowledged this important matter: “Given a specific interpretation of the [or a] theory, its consequences can be treated as predictions to be compared with data from the real world to determine the degree of correspondence between the theory and the world.” Reid Hastie and Gerald Stasser, Computer Simulation Methods for Social Psychology at 85 in Harry T. Reiss and Charles M. Judd, eds., HANDBOOK OF RESEARCH METHODS IN SOCIAL AND PERSONALITY PSYCHOLOGY (2000). The Hastie and Stasser discussion references a classic book, Morton Deutsch and Robert Krauss, THEORIES IN SOCIAL PSYCHOLOGY, 6-13 and 215-216 (1965).

⁷⁷ Brewer supra note 66 at 12-13; see also Aronson et al., supra note 66 at 134-35: “...the interplay between laboratory and field research is also critical to the development of an effective and applied social psychology. Basic process-oriented experimental research may isolate important causal processes; however, convincing demonstrations that those processes operate in applied settings are essential before theory can be converted into practice

⁷⁸ *Daubert v. Merrell Dow Pharmaceuticals* 509 U.S. 579 (1993); *Daubert v. Merrell Dow Pharmaceuticals* 43 F. 3d 1311 (1995). The *Daubert* decisions are addressed to factual proof rather than policy issues but the analogy is nevertheless close.

⁷⁹ *Daubert* (1993) at 591-92; *Daubert* (1995) at 1315.

Bendectin was a teratogen as epidemiological studies showing no established relationship with birth defects in humans.⁸⁰

The problem of the ecological validity of jury experiments has received considerable attention in the research literature⁸¹ and is not settled, but, at minimum, prudence should be exercised in recommending major policy changes bearing on juries from simulation experiments, especially when they are at variance with real world data.

IV. The Experiments on Punitive Damages: Major Problems of Ecological and Internal Validity

Although they ignore most real world data in deference to their experiments, the authors of *Punitive Damages* concede some problems of generalizability, acknowledging that simulation experiments cannot “replicate the sense of gravity and importance of a courtroom decision.”⁸²

They report that every effort was made to ensure realism: e.g. “Most of the experimental cases were based on real decided cases; jury instructions were taken directly from actual trials; and only citizens eligible to serve on real juries were sampled as research participants.”⁸³

In consequence, *Punitive Damages* addresses the problem of generalizing to the real world in the following summary:

An obviously central question is whether our results are reflected in behavior in the real world. An initial reason to believe that they are is that many of the tasks were quite realistic. As compared with real jurors, our mock jurors’ job was simplified, but it incorporated the basic responsibilities of the real world jurors. A second reason comes from the repeated replication of research results across experimental tests within the research program, and from replications in independent research programs.⁸⁴

It is important to look more closely at those experiments. A closer look reveals severe limitations on this claim. By first asking about ecological validity more than a few red flags about

⁸⁰ In *Daubert* (1995) at 962 the epidemiological data were considered to have a better fit than data derived well-constructed laboratory research. See Joseph Sanders, *BENDECTIN ON TRIAL: A STUDY OF MASS TORT LITIGATION* (1998) at Chapter 7, 175-192 for additional discussion.

⁸¹ See Shari Diamond, *Illuminations and Shadows from Jury Simulations*, 21 *LAW AND HUMAN BEHAVIOR* 561 (1997); Brian Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still out?* 23 *LAW AND HUMAN BEHAVIOR* 75 561 (1999).

⁸² *PUNITIVE DAMAGES* at 20.

⁸³ *Id.* at 19.

⁸⁴ *Id.* at 19-20.

making major policy recommendations from their findings appear. Additionally, the scrutiny of the experimental materials for ecological validity reveals problems that bear on internal validity and in one instance raises questions about the correspondence of findings across settings, that is external validity.

Real World Juries

In the introduction to *Punitive Damages* Professor George Priest reviews the basic procedural context in which real world juries decide punitive damages.⁸⁵ This context includes the voir dire process, opening statements, a trial involving evidence about liability and compensatory damages; and a decision about whether punitive damages are warranted. If punitive damages are deemed to be warranted, in a separate phase of the trial the jury hears evidence and arguments about the degree of misbehavior of the defendant and any remorse or other mitigating factors on the part of the defendant.⁸⁶ Finally, the jury receives instructions from the judge about how to decide those damages. Priest mentions in passing that it is also “increasingly common” for the plaintiff’s lawyer to make reference to the defendant’s wealth “though only in general terms.”⁸⁷

While this is a reasonable description of the basic components of the trial process, Professor Priest underplays the substance of what the jury hears at trial. Over a period of days or weeks the jury will be exposed to many documents and witnesses. This evidence will bear not only on the liability and compensatory damages but also on the behavior of the plaintiff and defendant. Jurors may see documents that reflect calculated behavior and reckless greed on the part of a defendant individual or organization, such as incriminating memoranda outlining deception or reckless behavior in the face of clear warnings. They may hear testimony from civilian or expert witnesses who corroborate the documents or supply additional details bearing on more than just

⁸⁵ Priest in PUNITIVE DAMAGES at 6-14.

⁸⁶ Statutes and case law vary from state to state as to a separate, or bi-furcated, trial. For example, in California, bi-furcation is mandatory on application of defendant. Jury will first determine liability/amount of compensatory damages & liability for punitive damages. Second phase reserved for determination of amount of punitive damages, which may not be mentioned during first phase. § 3295(d). In Florida bi-furcation is mandated on timely motion. At the first stage, the jury determines liability and the amount of compensatory damages and liability for punitive damages. If the jury finds punitives are warranted, it determines the amount of punishment, see *W.R. Grace & Co.-Conn. v. Waters*, 638 So. 2d 502, 506 (Fla. 1994). By contrast Alabama has a single stage trial; see *Life Ins. Co. of Ga. v. Johnson*, 701 So.2d 24, 532 (Ala. 1997), *overruling in part* 684 So.2d. 685 (Ala. 1996). In Arizona bi-furcation is not mandated. The court may order a separate trial of any claim or issue in furtherance of convenience or to avoid prejudice, or when separate trials will be conducive to expedition and economy. Ariz. R. Civ. P. 42(b).

⁸⁷ Priest, PUNITIVE DAMAGES at 13.

careless malfeasance. The jurors may hear first hand about the impact of this behavior on the victims. They may hear testimony bearing negatively on the credibility of plaintiff and defense witnesses or even see witnesses for either side exposed as liars during cross-examination. Jurors may also hear about the behavior of the defendant in previous incidents, in both the liability-compensatory phase and punitive phase of the trial. In the punitive phase they likely will hear arguments from defense counsel about the reasonableness of the defendant's behavior given what was known beforehand or about mitigating circumstances.⁸⁸ The jurors may hear expert evidence from both sides about the ability of the defendant to bear the economic burden of a punitive damage award. The defense may offer their own figures to rebut the plaintiff's suggested amount of punitive damages during final arguments, thus providing a counter-anchor to the plaintiff's *ad damnum*. The jury will be instructed to weigh all of this evidence with a plethora of knowledge gained during the first phase of the trial. Finally, the jurors, ranging between six and twelve in number, will be required to deliberate on the issues to reach a unanimous (or super-majority) verdict.

The above supplement to Professor Priest's summary thus reminds us that real world jurors do not begin the punitive phase of the trial with a blank slate about misbehavior. Nor does the punitive phase necessarily involve only the plaintiff presenting evidence and arguments; the defendant is allowed to present evidence and arguments. In short Priest's summary of the punitive damages process underplays the knowledge and education that the jury obtains during the trial.

When the jury receives its instructions from the judge on how to decide punitive damages it has a great deal of background information against which to interpret those instructions and reach a verdict. In Florida, for example, the jury is given instructions that it needs "clear and convincing evidence" that the conduct was so gross or flagrant as to show reckless disregard of human life or the rights of others or intentional misconduct or gross negligence to decide if punitive damages should be awarded against the defendant. In the instance of multiple defendants the jury is told that it may decide punitive damages are warranted against one defendant and not others or against more than one defendant.⁸⁹ In deciding the amount of punitive damages the jury is instructed to consider the "nature, extent, and degree of misconduct and related circumstances;" "each defendant's

⁸⁸ Richard Lempert also makes this point in a critique of one of the Exxon studies, see Richard Lempert, *Juries, Hindsight Bias, and Punitive Damage Awards: Failures of a Social Science Case for Change*, 48 DEPAUL LAW REVIEW 867, 877 (1999).

⁸⁹ The Florida Bar, STANDARD JURY INSTRUCTIONS IN CIVIL CASES, 2001 edition updated through 2002, Punitive Damages, Westlaw JICV FL CLE PD,

financial resources”⁹⁰ and “any other circumstances which may affect the amount of punitive damages.”⁹¹ The jury is also told that any disputed factual issues bearing on the amount to be assessed as punishment and deterrence to others should be determined by the greater weight of the evidence.⁹²

Since the authors of *Punitive Damages* make such broad policy recommendations we need to keep these real world facts firmly in mind and compare them to the experiments reported in *Punitive Damages* and prior published articles that describe the methodology of the experiments in greater detail than in the book. We need to ask questions about the degree to which the simulation experiments accurately mimicked the real world trial process and the degree to which the presentation of “trial” evidence and instructions was balanced. The first question speaks to ecological validity and the second speaks, in some important instances, to internal validity. Along the way some other important questions about validity can be raised.

From Outrage to Erratic Dollars in Eight-Sentence, Truncated “Trials”

Chapter 2 of *Punitive Damages* reports an experiment that attempted to study jurors’ ability to decide the dollar amounts of damages. The authors of that chapter conclude that the experiment demonstrates that jurors are erratic in translating moral judgments about defendant behavior into dollar awards.⁹³

Chapter 2 is based on a previously published law review article.⁹⁴ That article provides greater detail about the method and procedure of the experiment. The experiment was conducted in a “downtown motel” and “[m]ost [of the 899 “jury-eligible”] respondents completed their task in thirty to forty-five minutes. Each respondent first received three pages of general instructions on the law. Then, each respondent answered questions about a number of scenarios, ranging from “about one” to five, concerning an injury suffered by a plaintiff. The scenarios reported that compensatory damages had already been awarded. The ten personal injury scenarios used in the

⁹⁰ To be used only when financial resources are introduced, *id.*, note 3 re subparagraph (2) in PD 2nd

⁹¹ *Id.*.

⁹² *Id.*

⁹³ PUNITIVE DAMAGES at 42.

⁹⁴ Sunstein, et al., Assessing Punitive Damages (with Notes on Cognition and Valuation in Law), 107 YALE LAW JOURNAL 2071 (1998).

experiment ranged between six and ten sentences in length, with a median length of eight sentences.⁹⁵

After reading each scenario each mock juror was asked to respond to two six-point “bounded” rating scales measuring how much the defendant should be punished. The first scale was worded as follows: “Which of the following best expresses your opinion of the defendant’s action?” and the participants were provided with a six-point scale anchored at one end by “Completely Acceptable” and at the other by “Outrageous.” The second question was worded as follows: “In addition to paying compensatory damages, how much should the defendant be punished? Please express your opinion of the appropriate level of punishment.” The question was accompanied by a six-point scale anchored on the respective ends by “No Punishment” and “Extremely Severe Punishment.” The third question was an “open-ended” scale that asked, “In addition to paying compensatory damages, what amount of punitive damages (if any) should the defendant be required to pay as punishment and to deter the defendant and others from similar actions in the future? Please write your answer in the blank below.”⁹⁶

The respondents in the experiment did not have the rich evidential context of a real trial, had not decided upon liability for either compensatory damages nor punitive damages, did not have the benefit of arguments from plaintiff and defense counsel and did not have the advantage of testing their personal views against the views of others in a process of deliberation.

In addition to the paucity of information provided to the respondents, a second question arises about the scales. Do real juries make judgments about the appropriate level of punishment based on two questions that assess moral outrage on six-point scales? To what extent would those scale responses assess the egregiousness of the defendant’s behavior? Rating scales are commonly used in psychological research, but they have limitations and the best approach is to follow them with open-ended questions asking respondents to explain their reaction in their own words.⁹⁷ Would the simulating jurors have responded differently if, in addition, they had been asked to state their reasoning? Is it logical to compare responses on those bounded scales with the open-ended dollar scale? To be sure, as *Punitive Damages* points out, real world verdicts on punitives are made on open-ended scales, but real world juries do not assess moral reprehensibility on six-point

⁹⁵ Id. at 2147-2152.

⁹⁶ Id at 2152-2153.

⁹⁷ See Shari Diamond, Reference Guide on Survey Research, in Federal Judicial Center, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, 2ND ED (2000) at 251-260.

bounded scales and without open-ended discussion about the reprehensibility of the behavior or the specific or general deterrence that might be accomplished by an award of punitive damages. These considerations place limitations on the ability of the experimental findings to tell us about the real world.

This experiment needs to be evaluated along with a companion experiment reported in Chapter 3.

Deliberating About Dollars and Polarization Shifts in One-page, Plaintiff-biased “Trials”

Chapter 3 reports, in the authors’ words, the “results of a massive study of decisions by mock jurors”⁹⁸ Simulating jurors were presented with a summary of trial evidence and assigned to groups to deliberate on a punitive damages verdict. Most of the groups followed a simple majority rule in reaching their decision. More important, in comparison to the median of individual pre-deliberation judgments, dollar awards increased after group deliberation. Fully 27 percent of groups awarded dollar amounts that were as high or higher than the highest pre-deliberation judgment of any juror and the awards were less consistent and predictable than the mean or median amount awarded by the individual jurors. The authors of the experiment concluded that their findings are consistent with the “polarization shift” that has long been demonstrated in social psychology experiments whereby group discussion shifts pre-deliberation judgments toward extreme ends of psychometric scales.⁹⁹

Chapter 3 is also based on a previously published article that provides greater information about the experiment’s methodology.¹⁰⁰ In both the chapter and the article the authors make prominent the fact that that the study involved over 3000 jury-eligible respondents, and over 500 six-person deliberating juries. The authors further report that the individual jurors viewed a “video-taped narration” of one of a number of personal injury trials¹⁰¹ read by a professional actor, accompanied with a written script of the narration so that they could follow along. Then, approximately half of the subjects were asked to make a personal judgment on an eight-point scale

⁹⁸ PUNITIVE DAMAGES at 43.

⁹⁹ PUNITIVE DAMAGES at 31-32, 57-60, 61.

¹⁰⁰ Schkade, Sunstein and Kahneman, Deliberating About Dollars: The Severity Shift, 100 COLUMBIA LAW REVIEW 1139 (2000).

¹⁰¹ Fifteen “trial” scenarios, based on real cases were used in the research, see PUNITIVE DAMAGES, at 47.

(a “bounded” scale¹⁰²) of the degree to which the defendant should be punished¹⁰³ while the remainder of the respondents were asked instead to assess on an “open-ended scale”¹⁰⁴ an amount of damages that the defendant should be required to pay.¹⁰⁵

Next, the respondents were assigned to six-person groups and asked to reach a unanimous decision on the judgment they had just made individually, that is, either on the eight-point scale or the open-ended scale involving dollars. Next, they were handed a new form asking them to make an individual personal judgment on the other scale; that is, those who had received the bounded scale now responded to the open-ended scale and those who had responded to the open-ended scale now responded to the bounded scale. Finally, the respondents were again assigned to six-person groups and asked to deliberate to a unanimous decision. During deliberations they were required to reach their decision in thirty minutes. A total of 91 percent reached a unanimous decision on the punishment rating and 82 percent reached a unanimous decision on the open-ended dollar award judgment.¹⁰⁶

Let me be very clear here. The experiment used what is called a “within participants” design.¹⁰⁷ The procedure of having some subjects respond to one order of questions while another group of subjects responds in another order is known as counter-balancing for order effects and is an appropriate way of attempting to control biases that may result from making decisions in one order versus another and is used to attempt to increase internal validity of an experiment or survey.¹⁰⁸ “Within participants” experimental designs are commonly used in social psychology, but they are subject to problems bearing on internal as well as ecological validity. The internal validity problem is that by being required to make multiple decisions respondents may answer differently than if they were required to only make one decision.¹⁰⁹ As regards the problem of ecological validity, the essential point is that the conditions of the experiment do not mimic the conditions that lead to a single trial verdict on punitive damages that real world juries make.

¹⁰² See, Shari Diamond, *supra* note 97 at 229,251-264 (2000) for a discussion of closed and open-ended scales that are comparable to the terminology of “bounded” and “open-ended” scales.

¹⁰³ PUNITIVE DAMAGES at 45.

¹⁰⁴ Diamond, *supra* note 97.

¹⁰⁵ PUNITIVE DAMAGES at 45.

¹⁰⁶ PUNITIVE DAMAGES at 46.

¹⁰⁷ Brewer, *supra* note 66 at 22-23. Shadish et al., *supra* note 66 at 109.

¹⁰⁸ See, Shadish et al., *supra* note 66 at 115; Charles Judd, et al. RESEARCH METHODS IN SOCIAL RELATIONS, 6th Ed. (1991) at 92-98. There are other reasons for using “within participants” designs that need not be addressed in this review.

¹⁰⁹ See Brewer, *supra* note 66 at 22-23. The problem involves arcana of experimental procedure that are beyond the scope of this article but the reader is encouraged to refer to Brewer on this issue.

Neither the book chapter nor the article explicitly reports the length of the trial narration, but the appendix in the original article reports what can be assumed was a typical “trial” narration.¹¹⁰ It consisted of one page containing five short paragraphs, a combined total of thirteen sentences, about a products liability case. The first two paragraphs described the facts of the case: a child ingested medicine allegedly from a childproof safety cap and permanently weakened her respiratory system. In one sentence the subjects were told that the trial jury ordered the manufacturer to pay \$200,000 in compensatory damages. Another paragraph consisting of five sentences laid out the evidence about the manufacturer’s malfeasance. A two-sentence paragraph summarized the closing arguments by plaintiffs’ counsel and another two-sentence paragraph summarized the defendant’s closing arguments.

More than the obvious brevity of the stimulus “trial” deserves comment. The facts of the case are all biased in the plaintiff’s favor: the parents of the child were cautious; the overdose severely injured the child; the defendant’s childproof caps had a high failure rate and an internal memo not only downplayed the risks but said that the consequences of this misbehavior were trivial; there was a prior warning from the FDA.¹¹¹ The plaintiff’s closing emphasized these facts.¹¹² The two sentences relaying the final arguments of the defendant in the scenario said that the caps were generally effective and asserted that the FDA warning was “only communicated to them verbally.”¹¹³ In total, of the thirteen sentences comprising all of the facts and arguments in the trial, only two presented the defendant’s position, a plaintiff to defendant ratio of 5.5:1.¹¹⁴

Do real juries have such minimal facts and arguments upon which to make their judgments and in such a biased presentation? Are they told that another trial jury decided compensatory damages were warranted?¹¹⁵

Similar to the question raised about the experiment in Chapter 2 we need to ask: Do real juries make judgments about the appropriate level of punishment based on a simple instruction asking, “How much should the defendant be punished because of their actions and to deter the

¹¹⁰ Schkade et al., *supra* note 100 at 1174.

¹¹¹ *Id.* at 1174.

¹¹² *Id.*

¹¹³ *Id.*

¹¹⁴ Sentences favoring the plaintiff also included some long compound sentences so this is a conservative estimate of the plaintiff to defense ratio.

¹¹⁵ Obviously in some rare instances a different jury may be asked to decide only the punitive damages portion of the trial as when a defendant concedes liability and compensatory damages but contests the punitive damages or as when a punitive damages verdict is retried after a legal judgment vacating the first verdict, but these are exceptions.

defendant and others from similar action in the future?” by responding to an eight-point scale anchored at one end by “None” and at the other by “Extremely Severe”? Is it reasonable to compare responses on that bounded scale with the open-ended scale dollar scale? To what extent would that scale response, first individually and then through deliberation, affect responses to later dollar amounts compared to the discussion that real jurors might have about the egregiousness of the defendant’s behavior in their deliberations? As described above, some of the subjects were first directed to award an amount of damages without first giving considered attention to the egregiousness of the defendant’s behavior. But does that happen in real juries, especially under such minimal fact and argument conditions?

There is more. The authors of the research found that group deliberation among the simulating jurors tended to enhance the size of awards preferred by the individual jurors in the half hour allowed before deliberations were terminated. This led to the following conclusion:

We have found that...deliberation makes low-punishment judgments decrease and high-punishment judgments increase. It also makes –and this is our most important finding –dollar awards generally increase, while making high-dollar awards substantially increase, in a general severity shift.¹¹⁶

In the discussion of these findings¹¹⁷ the authors correctly report the extensive literature on the “polarization shift” that has shown that such shifts frequently occur in a variety of settings¹¹⁸ and that the phenomenon is likely to occur under two conditions: when the arguments of the various group members tend to be in the same direction and when they adjust their views to conform with a perceived social consensus.¹¹⁹

But now consider that, as described above, the facts and arguments in the thirteen- sentence trial summary were overwhelmingly in favor of the plaintiff. Under such conditions we can reasonably expect that most arguments and opinions generated by the simulating jurors would tilt toward the plaintiff and against the defendant. Putting it bluntly, the experiment did not provide a fair and balanced presentation of evidence.¹²⁰ A more balanced presentation might well have

¹¹⁶ PUNITIVE DAMAGES at 61.

¹¹⁷ PUNITIVE DAMAGES at 57 but reported in considerably more detail in the original article: see Schkade, et al supra note 100.

¹¹⁸ See, e.g. David Myers and Helmut Lamm, The Group Polarization Phenomenon, 83 PSYCHOLOGICAL BULLETIN 602 (1976) or the review of the polarization literature in Schkade et al. at ???

¹¹⁹ PUNITIVE DAMAGES at 58; Schkade et al, supra note 100.

¹²⁰ The importance of a fair and balanced presentation has been emphasized by Professor Hastie in an article discussing civil jury decision making. See Reid Hastie, *The Role of “Stories” in Civil Jury Judgments*, 32

obtained either no shift or more shifts toward lesser amounts of damages because the simulating jurors would have been able to consider off-setting arguments favoring the defendant.¹²¹

In summary, the gulf between the experimental simulation and the real world was wide, an ecological validity problem. In addition the trial summary presented to the simulating jurors contained no evidence favoring the defense. The imbalance in information provided to the simulating jurors involves a problem of internal validity because it does not allow us to rule out a

UNIVERSITY OF MICHIGAN JOURNAL OF LAW REFORM 227 (1999). Reviewing research on the story model the author referred to findings from experiments conducted on criminal jury decision making as follows:

For example, many criminal cases involve the presentation of only one story, by the prosecution, while the defense tactic is to 'raise reasonable doubts' by attacking the plausibility of that story. In these one-sided cases, jurors construct only one story, and confidence in the verdict is determined by coherence and fit of the single story to the verdict category. In this situation, a weak defense story is worse than no story at all; in fact, a weak prosecution story is bolstered and more guilty verdicts are rendered when a weak defense story is presented versus no defense story at all. (at 232, footnotes omitted from quotation.)

¹²¹ Strikingly, Professor Michael Saks, one of the authors of the Landsman et al. study of punitive damages, *supra* note __, suggests that data from that study seems consistent with my hypothesis on this matter. The experiment contained two versions of evidence, one of which involved weak evidence and the other moderate evidence supporting the plaintiff's position on liability. In a personal communication by email on February 18, 2004 Saks offered the following comments about the Landsman et al. experiment in response to my hypothesis:

The two levels of evidence strength for compensatory liability were pretested to be weak and moderate. In the moderate strength case, the plaintiff had a good bit of exposure to the defendant's products and smoked barely at all. In the weak case, the plaintiff had less verifiable exposure to the defendant's products and was a smoker (so the defendant's compensatory liability was less clear). The evidence on punitive liability was identical in all conditions. (Our "beryllium" case was modeled after asbestos: the defendants clearly knew and actively concealed the fatal harm of their product.)

To the jurors' credit, the strength of the compensatory liability case had no effect on either the likelihood of finding punitive liability (see Table 6) or the amount of punitive damages awarded (see Table 7). As the compensatory liability of the defendant became more clear, they did not become more likely to find punitive liability or to award more in punitive damages.

By comparing the damages amounts for certain conditions in Tables 7 and 8, one can see that it is not apparent that the "amplification" effect hypothesized by Sunstein et al. occurred. That conclusion is even more clear from looking at the "expected awards" in the decision trees for individual jurors versus deliberating juries in Figures 1 and 2 at the very end of the article. (Focusing on unitary trials, juries can be expected to award less in damages than nondeliberating individual jurors.)

In addition consider an experiment by Martin Kaplan and Charles Miller, *Judgments and Group Discussion: Effect of Presentation and Memory Factors on Polarization*, 40 *SOCIOMETRY* 337 (1997). Some simulating juries listened to a tape-recording of facts incriminating a criminal defendant while others heard exonerating facts. Some jurors heard the facts in the same order (homogeneous order) and others heard them in a different order (heterogeneous order). The experiment found that, as hypothesized, a greater number of facts were remembered and shared by the juries under the heterogeneous compared to the homogeneous conditions. Moreover, in the exonerating condition there were group polarization shifts away from a guilt decision and in the incriminating conditions group discussion produced polarization shifts toward guilt. Although conducted in the context of a criminal trial the polarization shifts in this experiment are consistent with my suggestion that a more balanced presentation of facts or facts tending to favor the defendant in the Chapter 4 experiments would have produced no shifts or shifts away from larger punitive awards.

plausible alternative causal explanation of the results, namely that rather than a bias on the part of the jurors the polarization shift was caused by the imbalance of information overwhelmingly favoring the plaintiff. Viewed from this perspective, generalizations about the effect of jury deliberations on deciding damages are hazardous.¹²² The conclusions of “higher than expected” and “greater than appropriate” post-deliberation awards in the study may well be a result of the atypical conditions and biased presentation under which the subjects operated, notwithstanding the fact that the experiment contained over 3000 subjects.

The Ad Damnum, Simulated Anchoring Effects and AWOL Defense Lawyers

Chapter 4 of *Punitive Damages* presents two experiments purporting to assess the impact of the plaintiff’s *ad damnum* on punitive awards.¹²³ The experiments were designed to test the hypothesis that the unbounded task of making dollar awards was likely subject to psychological anchoring produced by the plaintiff’s *ad damnum*.¹²⁴ Chapter 4 concludes that there was a large effect of the size of the plaintiff’s reward request; the more that was requested the more that was awarded.¹²⁵

Once again consider the methodology of the experiments.¹²⁶ The simulating jurors were told that liability for the accident (an environmental spill of a toxic chemical) had been established in a prior legal proceeding and the defendant had paid \$24.5 million in compensatory damages and fines. They were further told that the proceeding had determined that the company’s actions

¹²² Although I focus here on the polarization shift, Wissler, Saks and Hart, Decisionmaking about General Damages: A Comparison of Jurors Judges and Lawyers, 98 MICHIGAN LAW REVIEW 751, 802, note 147, point out that the claim of greater variability made by Sunstein et al. confuses the polarization phenomenon with the mass of cases dealing with punitive damages.

¹²³ Hastie, Schkade and Payne, Do Plaintiffs’ Requests and Plaintiffs Identities Matter?, Chapter 4 of PUNITIVE DAMAGES. The chapter also investigates whether the residence of plaintiffs and defendants (a local entity or out-of-state) would have an effect on awards. Statistical analyses produced equivocal results regarding residence location and will be ignored here.

¹²⁴ Chapter 4 also reports that the first of the experiments also attempted to assess whether out-of-state defendants, as opposed to local defendants are treated more harshly. I do not deal further with this part of the experiment except to note that in summarizing their findings the authors, asserted, id. at 73, that that “...punitive damages amounts were influenced by the plaintiff’s location,” and followed followed immediately with a contradictory sentence reporting that “...we did not find statistically reliable effects due to defendant location.” Statistically, it is improper to make a claim of a finding when the statistical tests do not support the hypothesis. Sometimes authors report a “trend” toward a certain hypothesis when the data are in the predicted direction of a hypothesis, but no qualification exists in the authors’ summary of findings.

¹²⁵ Id. at 73.

¹²⁶ Chapter 4 is based on Reid Hastie et al., Juror Judgments in Civil Cases: Effects of Plaintiff’s Requests and Plaintiff’s Identity on Punitive Damage Awards, 23 LAW AND HUMAN BEHAVIOR 445 (1999).

constituted reckless conduct and that punitive damages were warranted.¹²⁷ They then viewed a videotaped narration, accompanied by a written summary and instructions. The presentation lasted 35 minutes.¹²⁸ The award-anchor conditions were created by varying the plaintiff's closing argument, ranging from low to high amounts of suggested damages.¹²⁹ Aside from the artificiality of not being involved in the liability and compensatory stages of the trial or having the advantages of deliberation, the subjects heard not a word from the defense side about why the punitive damages should not be so high, nor were they provided with a defense counter-anchor or expert evidence. (If the case had been a real trial it would appear that the defendant corporation had solid grounds to sue its counsel for legal malpractice.) In short, compared to real trials, the experiment presented only the plaintiff's side of the argument.

Punitive Damages' assertion that psychological anchoring has been demonstrated in many experiments in the field of psychology¹³⁰ is only partly accurate, especially when other experiments involving simulating jurors deciding compensatory awards are considered. This is an important matter because Professor Sharkey, critical of other parts of *Punitive Damages*, has taken the anchoring effect at face value, concluding, that "[behavioral research has demonstrated that the anchors selected have a disproportionate influence on the outcome of the decision]"¹³¹ and, further, that "Sunstein et al. demonstrate that anchoring effects wreak havoc in the punitive damages realm."¹³² Other law-trained scholars may also accept this conclusion. It is misleading.

Anchor effects are far more complicated than the impression one obtains from reading Chapter 4, especially when it comes to experimental studies of jury decision-making. Consider an experiment by Saks et al.¹³³ In one condition simulating jurors were provided both a low and a high award amount as anchors. The size of compensatory awards did not differ from a condition providing no anchors, but the variability of the awards was reduced. Marti and Wissler followed

¹²⁷ Id. at 65.

¹²⁸ Id. at 65-66, .

¹²⁹ Id. at 66 and 71.

¹³⁰ The authors, at 63, note 4, cite a review by Payne, Behavioral Decision Research; An Overview, in Michael Birnbaum et al., MEASUREMENT, JUDGMENT AND DECISION MAKING (1998) as the basis for their rationale, along with a study by Bornstein and Chapman, The More you Ask For, the More You Get: Anchoring in Personal Injury Verdicts, 10 APPLIED COGNITIVE PSYCHOLOGY 519 (1996). Anchoring effects have been demonstrated in many settings, see, e.g. Amos Tversky and Daniel Kahneman, Judgment Under Uncertainty : Heuristics and Biases, 185 SCIENCE 1124 (1974); Jackowitz and Kahneman, Measures of Anchoring in Estimation Tasks, 21 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 1161 (1995).

¹³¹ Sharkey, supra note 17 at 408 (footnote omitted).

¹³² Id.

¹³³ Michael Saks et al. Reducing Variability in Civil Jury Awards, 21 LAW AND HUMAN BEHAVIOR 243 (1997).

up on this finding in an experiment that varied the amounts that both the plaintiff and the defense suggested for pain and suffering.¹³⁴ The size of the awards and the variability of those awards was influenced by both the plaintiff request and the defense request.¹³⁵

Professors Edie Greene and Brian Bornstein's book, *Determining Damages*, devoted to experimental studies of how juries decide damages, thoroughly reviewed anchoring research as it relates to jury decisions.¹³⁶ Their review considered simulation studies in which the defense offered a counter-figure to that of the plaintiff, in which expert economic evidence was introduced, and in which the simulating jurors deliberated. The findings showed that while simulating jurors were influenced by anchors,¹³⁷ the effects were much more nuanced than the *Punitive Damages* experiments and their authors suggest: to wit, defense counter-figures had moderating effects. In an experiment involving deliberations Greene and Bornstein concluded that "[a]pparently jury awards may be reduced when the defense presents a counter-offer, at least when that amount is offered by an expert."¹³⁸ In other words anchoring effects change when defense evidence and arguments are introduced. To repeat, such counter-anchors were absent from the *Punitive Damages* experiment.

Thus, the simulation studies reported in Chapter 4 not only lack ecological validity, by among other things, failing to create conditions of evidence and counter-proposals by defendants, but they also lack external validity when compared to other experimental studies of jury award behavior. In addition the fact that the materials provided the respondents were heavily weighted in favor of the plaintiff, similar to the "Deciding Dollars" experiment in Chapter 3, raises problems of internal validity.

In an article published in the *Stanford Law Review* subsequent to *Punitive Damages* Sunstein, Kahneman, Schkade and Ritov acknowledged that "... *judgment shifts are easy to generate in experimental settings*"¹³⁹ (italics added.) Yes, and perhaps we may conclude from

¹³⁴ Mollie Martie and Roselle Wissler, Be Careful What you Ask For: The Effects of Anchors on Personal Injury Damages Awards, 6 JOURNAL OF EXPERIMENTAL PSYCHOLOGY: APPLIED 91 (2000).

¹³⁵ Id. Award size and variability increased with the amount requested by the plaintiff, except when the amount suggested was extreme, in which event the amount decreased. The defense suggestion had a mirror effect: Award size and variability decreased in response to a defense rebuttal, except when the amount suggested was very low, in which case awards increased.

¹³⁶ Edith Greene and Brian Bornstein, DETERMINING DAMAGES: THE PSYCHOLOGY OF JURY AWARDS (2002) at 150-156.

¹³⁷ Id. at 155, 156

¹³⁸ Id. at 155.

¹³⁹ Sunstein et al., Predictably Incoherent Judgments, 54 STANFORD LAW REVIEW 1153,1203 (2002).

Chapter 4 that plaintiff anchors do produce judgment shifts when experiments provide only one side of the evidence and arguments. But real juries operate from a very different context than that provided in the experiments reported in Chapter 4. The research reported in Chapter 4 presents very major ecological, external and internal validity problems.

A Study Involving Legal Rather Than Fact Decisions and One-sided Presentation of Evidence

Chapter 5 describes an experiment ostensibly testing whether individual jurors and deliberating juries can accurately assess whether liability for punitive damages is warranted. The respondents in the experiment were provided with four case synopses and asked to determine if the cases met the legal requirements for liability. The deliberations of the simulating jurors were videotaped and subsequently analyzed. In three of the four cases the majority of the jurors decided in favor of the plaintiff, suggesting, the authors proclaim, a pro-plaintiff tilt.¹⁴⁰ The Chapter concludes, in part, that “the tendency to find the defendant liable was partly due to juries’ failure to systematically consider the full set of legally necessary conditions of the verdicts they rendered.”¹⁴¹ It further concludes that jurors displayed low levels of comprehension and memory for information conveyed in legal instructions.¹⁴² The authors make the following assertion: “We believe that the essential finding, disturbingly poor comprehension, is a valid description of what occurs in actual jury trials, although this conclusion is based on our mock-jury simulation, which is far from courtroom conditions.”¹⁴³

¹⁴⁰ In Chapter 5 this assumption is not made explicitly but in the original article upon which the chapter is based, see Hastie et al, *A Study of Juror and Jury Judgments in Civil Cases: Deciding Liability for Punitive Damages*, 22 *LAW AND HUMAN BEHAVIOR* 287 (1998), the authors made a number of statements implying anti-defendant attitudes on the part of juries. At the beginning of their article they claimed: “What is incontrovertible is that in a nonnegligible number of cases juries render distinctly harsh antidefendant judgments and set dramatically large awards,” at 288. In interpreting their results the authors concluded that “In the cases we studied individual jurors exhibited a persistent tendency to favor plaintiffs, concluding that punitive damages were warranted when judges had concluded they were not,” at 306. The article further speculated that “possibly sympathy for injured plaintiffs led them to ignore essential legal considerations,” at 307 and “Furthermore, if as we have suggested, there is a bias in jury decisions against unsympathetic defendants and in favor of injured plaintiffs, increasing the requisite standard of proof should help reduce the rate of improper verdicts,” at 308. In a rejoinder and critique of that article, Vidmar, *Juries Don’t Make Legal Decisions! And Other Problems: A Critique of Hastie et al. on Punitive Damages*, 23 *LAW AND HUMAN BEHAVIOR* 705, at 712 (1999), I pointed out that the presumption that jury awards favoring plaintiffs do not allow a conclusion of anti-defendant biases as the authors presumed, but merely that the decisions were against the defendants.

¹⁴¹ *PUNITIVE DAMAGES* at 77.

¹⁴² *Id.* at 92.

¹⁴³ *Id.* at 92.

How far was the simulation from courtroom conditions? Chapter 5 is based on an article that was published previously.¹⁴⁴ In an article published in 1999 I offered a number of serious grounds of criticism.¹⁴⁵

Most important, I pointed out that there is a fatal conceptual flaw bearing on legal relevance in the underlying assumption of the experiment. The simulating jurors were asked to make decisions about whether punitive damages should be allowed. However, the decisions that the subjects were asked to make were decisions that fell within the province of the trial judge, not the jury.¹⁴⁶ The case synopses provided to the simulating jurors were ones in which the decision of the trial judge to allow or not allow punitive damages was appealed. In each case an appeals court decided that the trial judge, as a gatekeeper on legal issues, should not have permitted the jury to consider punitive damages. Juries make fact decisions, not legal decisions. As a consequence, the experiment lacks legal relevance.¹⁴⁷

While the above flaw in legal reasoning underlying the experiment is sufficient grounds for dismissing Chapter 5 in its entirety, additional problems with the research are worth pointing out. The experiment created artificial and truncated conditions somewhat similar to the experiments discussed above.¹⁴⁸ The jurors were told that liability and compensatory damages had already been

¹⁴⁴ Hastie et al, A Study of Juror and Jury Judgments in Civil Cases: Deciding Liability for Punitive Damages, 22 LAW AND HUMAN BEHAVIOR 287 (1998).

¹⁴⁵ Vidmar, supra note 11. Additional debate and a reply by the authors Hastie et al. accompanied that article: Hastie et al., Reply to Vidmar, 23 LAW AND HUMAN BEHAVIOR 715(1999); Phoebe Ellsworth, Sticks and Stones, 23 LAW AND HUMAN BEHAVIOR 719(1999); Robert MacCoun, Epistemological Dilemmas in the Assessment of Legal Decision Making, 23 LAW AND HUMAN BEHAVIOR 723 (1999).

¹⁴⁶ Vidmar supra note 11.

¹⁴⁷ Id. at 706-710. The authors of Chapter 5 claim that reliance on the court decision in those cases is not the test of juries decisions but that was not always so, see Vidmar's critique of the original article, supra note 11 at 707-710. Moreover, Professor Feigenson's review, supra note 21, has pointed out that the authors of Chapter 5 do seem to intend that the reader infer that the discrepancies between their simulating jurors responses and the actual judicial outcomes are the appropriate standard. Feigenson at 246, observed:

"...the authors [of Punitive Damages] do seem to intend that readers infer from the discrepancies between experimental and actual outcomes that the mock juries that voted for punitives erred. They write that they chose the four cases 'because they are frequently cited as precedents, and because the *proper action* on the issue of punitive damages had been decided as a matter of law by trial or appellate court review of the original proceedings' (p.80)(emphasis added by Feigenson), implying that the actual cases create a reliable benchmark after all. More directly, they write that 'discussion of issues that were legally relevant reduced the tendency to decide, *incorrectly*, that the punitive damages were warranted" (p.90) (emphasis added by Feigenson).

Moreover, as I pointed out in my original critique of that research, Vidmar, supra note 11 at 709-710, in two of the cases used in the experiment the trial and appellate judges struggled over the issue of whether, under the law, punitive damages were permissible and were not unanimous in their conclusions; and in the other two cases the issues were close enough to encourage plaintiff lawyers to file an appeal against the trial judge's decision to disallow punitives.

¹⁴⁸ Vidmar, supra note 11.

decided and they were only to decide if the evidence supported a finding that punitive damages were warranted.¹⁴⁹ Thus, there is again the problem of ecological validity because the simulating jurors did not hear or decide the first part of the case and had a very minimalist version of the facts and arguments. In addition there are major problems that compromise the experiment's internal validity.

The four cases selected to present to the simulating jurors provided only instances in which the appeal courts concluded that punitives were impermissible under the law. The experiment did not provide the jurors with comparable cases in which appeals courts determined that punitive damages were warranted. Would jurors provided with these alternative cases have decided punitive damages were not warranted? In short an experiment testing whether juries would decide the law in the same way as judges would have provided juries with "permissible" cases as well as "impermissible" cases and allowed a test of whether the jurors' decisions showed a pro-plaintiff/anti-defendant tilt.¹⁵⁰ Of course, from a practical standpoint devising such an experiment would only serve to satisfy appropriate experimental rigor since the experiment would still be legally irrelevant.

There is still more. The simulating jurors also deliberated as well as rendered individual verdicts. The authors concluded that deliberations produced a tilt in favor of finding that punitive damages were warranted. The critique which I made of the original article drew attention to the fact that by including "hung juries," that is, those who could not reach unanimous verdict in the allotted time, the researchers inflated the number of juries favoring punitive damages.¹⁵¹ Since the burden of proof is on the plaintiff, hung juries should count as favoring no punitives. When the data are adjusted for this inappropriate counting, the number of juries favoring punitive damages was only 58 percent. From this insight one could make a plausible argument that, on average, the experimental juries tended to be relatively comparable to the various judges involved in these cases who did not garner consensus on the question of whether punitive damages were warranted.¹⁵²

In reviewing that study again I now offer a further criticism. Similar to the experiments reported in Chapters 2 and 3, the trial materials, at least as reflected in the example provided in

¹⁴⁹ Hastie et al, *supra* note 138.

¹⁵⁰ Vidmar, *supra* note 11 at 710.

¹⁵¹ Vidmar, *supra* note 11 at 710.

¹⁵² See discussion in note 146 *supra*, and Vidmar, *supra* note 11 at 710.

Chapter 5 and in the original article upon which Chapter 5 is based, was heavily biased in favor of the plaintiff. That case summary, *Jardel Company, Inc. v. K. Hughes*,¹⁵³ consisted of 28 sentences.¹⁵⁴ All of those sentences presented the plaintiff's side of the story. Two additional sentences presented the plaintiff's arguments in favor of punitive damages.¹⁵⁵ The defendant's arguments against punitive damages consisted of three sentences.¹⁵⁶ In total, then, the case summary consisted of 30 sentences favoring the plaintiff and three sentences presenting the defendant's side of the case, or a 10:1 ratio of plaintiff information to defendant information.¹⁵⁷ Thus, again in this experiment, we see that the stimulus materials were heavily slanted toward a finding for the plaintiff. Perhaps such an imbalance of information sometimes occurs in real world punitive damages cases, but the imbalance in the evidence presented to the jurors is unremarked by the authors of *Punitive Damages*¹⁵⁸ although it seems obvious that such an imbalance of information could--and I submit, probably does-- explain the pro-plaintiff tilt in the data.

In summary, there are major problems in making inferences from the experiments reported in Chapter 5 that involve legal relevance, ecological validity and internal validity.

Mandated Deterrence and Risk Analysis Experiments Using "Fourteen-Sentence" Trials

The authors of the research in punitive damages appear to be committed to a law and economics view that the purpose of punitive damages should be primarily (perhaps exclusively) to deter behavior, rather than also considering the retributive or moral role that punitive damages may serve. For example, in a reply to a critique of one of their earlier experiments Professors Hastie and Viscusi stated: "We have described the juror as a risk manager because we believe that the most useful framework within which to understand the punitive damages decision is to view it as a governing mechanism that is aimed to deter or control behavior that reduces the general social

¹⁵³ PUNITIVE DAMAGES at 93-95. The original case is *Jardel Company, Inc. v. K. Hughes*, Del. Supr., 523A 2d 518 (1987).

¹⁵⁴ I count the sentences beginning with "The specifics of the case" on page 93 and end with the sentence just before the simulating jurors are told that plaintiff Hughes was compensated for her expenses on page 94.

¹⁵⁵ PUNITIVE DAMAGES at 94.

¹⁵⁶ *Id.* at 95.

¹⁵⁷ The jury instructions involved 19 sentences, see Hastie et al., *supra* note 142 at 310-311.

¹⁵⁸ The authors do not present the other case summaries in their published research and I have not obtained them to ascertain if they are the same length but Hastie et al., *supra* note 141 at 291, indicate that summaries ranged between 1000 to 1500 words in length,

welfare.”¹⁵⁹ Although this view is qualified somewhat in *Punitive Damages*¹⁶⁰ it remains a significant theme in the book.

Three chapters in *Punitive Damages* report experiments assessing whether respondents can appropriately respond to instructions that bear on the probability of deterrence.¹⁶¹ In one experiment reported in Chapter 8 samples of jury-eligible citizens and samples of law students were given brief case summaries that varied the probability of a defendant’s tortious behavior being detected.¹⁶² Following case synopses involving an injured person, respondents in the experiment were given the following instructions:

In situations like this, the victims who deserve compensation do not always receive it because (1) they don’t know what caused their problem and therefore don’t sue, (2) they don’t know that they can sue, or (3) they sue and lose, even though they deserve to win under the law , because their lawyers are not good enough. Research has shown that in only 1 out of 100 situations where someone has an experience like Joan Glover is the company eventually required to pay compensation to the victim.¹⁶³

Each respondent judged three cases, each of which varied the probability of detection.¹⁶⁴ Although there was a trend for the likelihood of detection to affect dollar judgments it was not statistically significant and was in a direction opposite to the predicted direction,¹⁶⁵ causing the authors to conclude that changes in the probability of detection did not affect the amounts awarded.

From the reported instructions it appears that the respondents in the experiment were not asked to assume the role of jurors, but rather to merely indicate their awards. This poses a serious problem. In a critique of an earlier experiment in the Exxon corpus of research, also

¹⁵⁹ Reid Hastie and W. Kip Viscusi, *Juries, Hindsight, and Punitive Damages Awards: Reply to Richard Lempert*, 51 DEPAUL LAW REVIEW 983, 994 (2002). They further added that in response to Professor Lempert’s view and that of others that part of the reason is that , “...we are unable to spell out a clear rationale for the moral role of the jury or a method to evaluate the moral quality of jurors’ judgments, at 995.

¹⁶⁰ PUNITIVE DAMAGES at 109: “These ideas [optimal deterrence being the primary aim of punitive damages] are controversial and we do not as a group, intend to take a position on them here.”

¹⁶¹ PUNITIVE DAMAGES, Chapters 7, 8 and 9.

¹⁶² *Id.* at 136.

¹⁶³ *Id.*

¹⁶⁴ The design was a randomized repeated measures design, commonly used in psychological research. Repeated measures designs are appropriate ways of assessing the effects of various conditions but the process of decision-making may be influenced by the respondent knowing that more than one decision has to be made. This is not necessarily a fatal flaw in the design but in this case note that jurors decide only one case, not several. See note ^{supra} and accompanying text.

¹⁶⁵ PUNITIVE DAMAGES at 137: “ To be sure, there does appear to be a positive trend in the means, but this difference is not statistically significant –and even if it were, it is in the opposite direction to the deterrence argument....”

dealing with judgments of risk,¹⁶⁶ Richard Lempert obtained the questionnaires and raw data from the experiment.¹⁶⁷ He noted a major methodological flaw affecting the experiment's internal validity, namely that rather than receiving identical instructions respondents in one condition were told to respond as ordinary citizens while in another they were told to assume the role of jurors.¹⁶⁸ Lempert's reanalysis of the data showed substantial differences between jurors assigned to the role of jurors compared to persons who received the same instructions who were not assigned to the role of jurors.¹⁶⁹ While we cannot say whether or how the instructions in Experiment 1 might have produced different effects if the respondents had been instructed to take the role of jurors, Lempert's finding raises serious questions about both its external and ecological validity.¹⁷⁰

A second "experiment" in Chapter 8 involved a survey of law students at the University of Chicago.¹⁷¹ They were asked if about the degree to which they agreed with the optimal deterrence approach to punitive damages by responding to a scenario involving reckless employer safety practices. The students were not asked to assume the role of jurors. Strikingly, very strong majorities of the law students rejected a judicial ruling based on optimal deterrence, presumably because they believed "invidious behavior must be punished."¹⁷²

The results of this second experiment bear on the results of research reported in Chapter 9 that asked respondents to apply the Polinsky-Shavell mathematical formula based on optimal deterrence.¹⁷³ Polinsky and Shavell set forth a mathematical formula based on optimal deterrence, namely that punishment levels should be based on the reciprocal of the probability of detection of negligent behavior.¹⁷⁴ Those authors proposed model jury instructions using this formula. In the research reported in Chapter 9 samples of jury eligible adults were exposed to variations of a summary of an environmental accident case that varied the probability that the defendant's tortious

¹⁶⁶ Reid Hastie et al., *Juror Judgment in Civil Cases: Hindsight Effects on Liability Judgments*, CRJP Technical Report #376 (1998). Data from this experiment formed part of the data base cited in Hastie and Viscusi, *What Juries Can't Do Well: The Jury's performance as a Risk Manager*, 40 ARIZONA LAW REVIEW 901 (1988).

¹⁶⁷ Richard Lempert, *Juries, Hindsight Bias, and Punitive Damage Awards: Failures of a Social Science Case for Change*, 48 DEPAUL LAW REVIEW 825 867 (1999).

¹⁶⁸ *Id.* at 882-883.

¹⁶⁹ *Id.*

¹⁷⁰ It is also noteworthy that in analyzing other parts of the Hastie and Viscusi study, *supra* note 163, Lempert also showed that their data did not demonstrate that judges are less susceptible to hindsight biases than jurors, see Lempert, *supra* note 88 at 883.

¹⁷¹ PUNITIVE DAMAGES at Chapter 8, page 140.

¹⁷² *Id.* at 140.

¹⁷³ The research reported in Chapter 9 was originally published in W. Kip Viscusi, *Corporate Risk Analysis: A Reckless Act?* 52 STANFORD LAW REVIEW 547 (2000).

¹⁷⁴ A. Mitchell Polinsky and Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARVARD LAW REVIEW 869 (1998).

behavior would be detected.¹⁷⁵ The respondents were recruited for an opinion survey and were not allowed the option of considering the degree to which the behavior of the defendant deserved retributive punishment. The respondents were then presented with three personal injury case synopses. The longest of these three synopses consisted of fourteen sentences. One of the fourteen sentences told the jurors that a jury had already determined that the defendant was liable and had assessed compensatory damages. Without further instructions they were asked to specify the amount of punitive damages that they believed were appropriate. No other information was given about the rich factual context out of which the disputes arose, nor did the respondents have an opportunity to compare their individual views with those of other respondents.

The respondents did not perform well in applying the Polinsky-Shavell hypothesis. It is not clear what implications this study reported in Chapter 9 has for real world juries. Aside from the limitations of the stimulus materials used for the trial the Polinsky-Shavell instructions do not correspond with the instructions that real world jurors are provided, there are additional problems. The first is that deterrence is only one goal of punitive damages and by artificially limiting the respondents' decisions of optimal deterrence the experiment lacked ecological validity. Second, as Professor Feigenson has pointed out, the Polinsky-Shavell instructions are so complicated that we should not expect laypersons to be able to apply them readily.¹⁷⁶

Indeed, Chapter 9 reports that “[a]lthough college graduates and those with professional degrees often did the survey in a manner that followed the Polinsky-Shavell instructions, a considerable portion of this group did not carry out these instructions.”¹⁷⁷ The chapter further notes that [t]he experiment did not distinguish whether people were unable to implement these instructions or were unwilling to follow these instructions.”¹⁷⁸ The second experiment of Chapter 8 suggests that at least some of the problem with the experiment in Chapter 9 may have been related to basic beliefs that invidious conduct deserves retributive punishment. These beliefs may have interfered with any calculations the respondents were asked to make or perhaps they were just unwilling to apply the deterrence formula.¹⁷⁹

¹⁷⁵ The research was previously published in Sunstein et al. Do People Want Optimal Deterrence?, 29 J. LEGAL STUDIES. 237 (2001).

¹⁷⁶ Feigenson, *supra*, note 21 at 272-276.

¹⁷⁷ PUNITIVE DAMAGES at 162. Note that the students were from the University of Chicago, an institution noted for its emphasis on law and economics.

¹⁷⁸ *Id* at 163.

¹⁷⁹ Feigenson, *supra* note 21 at 271-276 makes this same point in much greater detail.

It is noteworthy that other jury decision-making research indicates that when trial judges as well as jurors are asked to apply probabilistic formulas in abstract settings they tend to perform poorly.¹⁸⁰ On the other hand bodies of research indicate that when decisions are made in the context of concrete examples and settings, lay persons can make relatively accurate probabilistic judgments.¹⁸¹ Given the complexity and artificiality of the research reported in Chapter 9 as well as the fact that the stimulus materials do not correspond with actual conditions under which real juries operate or with legal doctrine providing for retributive as well as deterrent purposes of punishment the application of these findings to legal policy are difficult to discern.

Summarizing: Experimental Simulations and Experimental Validity in Perspective

The conclusion to *Punitive Damages* refers to the twenty empirical studies reported in the volume and asserts that “[t]hese experimental tests were conducted on thousands of jury-eligible citizens, on hundreds of mock juries, and on a few hundred experienced trial judges asked to render punitive damages verdicts under legally realistic task conditions.”¹⁸² By focusing on the actual methodological details of the experiments we can see that more than a few questions need to be raised about the meaning of “legally realistic.” The gap between the experiments and the conditions under which real world juries operate is wide and deep. Moreover, in some instances the experiments presented such one-sided presentations of facts that serious questions can be raised about the causal inferences the authors drew from their data.

The artificiality of experimental jury simulations has raised many questions about the extent to which generalizations can be made to actual jury verdicts.¹⁸³ In a comprehensive review of jury research methodology Robert MacCoun pointed out that while simulation experiments can shed light on how variables probably influence verdicts, [w]hen the objective is to estimate precisely the magnitude of trends and patterns in actual jury trials, ...archival analysis [such as

¹⁸⁰ For a review of these studies see Neil Vidmar and Shari Diamond, *Juries and Expert Evidence*, 66 BROOKLYN LAW REVIEW 1121 (2001) at 1134-1139, 1149-1158

¹⁸¹ *Id.*

¹⁸² PUNITIVE DAMAGES at 211.

¹⁸³ See 3 LAW AND HUMAN BEHAVIOR, combined issue 1/2 (1979); Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?* 23 LAW AND HUMAN BEHAVIOR 75 (1999); Diamond, *Illuminations and Shadows from Jury Simulation*, 21 LAW AND HUMAN BEHAVIOR 561 (1997)..

those “real world” studies discussed in the first section of this article] is more appropriate. The role of mock- jury experimentation is to explain those patterns.”¹⁸⁴

It is important to re-emphasize that my critique of *Punitive Damages* should not be taken as a general indictment of experimental simulation research on juries. Beginning in the 1970s researchers began conducting experiments on juries.¹⁸⁵ Many of these experiments used trial stimulus materials that were as spare in detail as the ones used in *Punitive Damages*. However, a second generation of experiments quickly arose and became more sophisticated in taking cognizance of legal policy matters and the richness of the jury’s factual environment, including judicial constraints on juries. Researchers developed theories about jury behavior, tied those theories to broader bodies of psychological research and checked their results against real world studies. Ironically, a series of jury simulation studies by one of the authors of *Punitive Damages* resulted in the “story model” of juror decision making processes¹⁸⁶ that supplanted earlier models of jury decision making and is, unquestionably, not only the prevailing theory today, but one that garners considerable support from real world data as well as from many other experiments, some using minimalist trial materials and others using much more sophisticated jury simulations. Taken together the incremental knowledge gained from the large body of simulation studies does provide a basis for assisting legal policy makers.¹⁸⁷

One problem (though I hasten to add, not the only problem) with *Punitive Damages* is the lack of follow-through experiments that assess plausible alternative hypotheses.¹⁸⁸ Consider the *ad damnum* experiment reported in Chapter 4. Putting aside the biased facts favoring plaintiffs in the stimulus materials, it would be reasonable to conduct a first experiment looking only at the effect of the plaintiff’s *ad damnum* request because the researcher might want to simply isolate only the effects of the plaintiff’s request. However, careful researchers would have followed up with additional experiments seeking to find out about the potentially countervailing effects of

¹⁸⁴ Robert MacCoun, Inside the Black Box: What Empirical Research Tells Us About Decisionmaking By Civil Juries. In Robert Litan, ed., VERDICT : ASSESSING THE CIVIL JURY SYSTEM (1993).

¹⁸⁵ I See generally, Devine et al., Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups 7 PSYCHOLOGY, PUBLIC POLICY AND LAW 622 (2001), Edie Greene and Brian Bornstein, DETERMINING DAMAGES: THE PSYCHOLOGY OF JURY AWARDS (2003); Neal Feigenson, LEGAL BLAME: HOW JURORS THINK AND TALK ABOUT ACCIDENTS (2000); Valerie Hans, BUSINESS ON TRIAL (2000).

¹⁸⁶ Reid Hastie, Stephen Penrod and Nancy Pennington, Inside the Jury (1983); Pennington and Hastie: A Cognitive Theory of Juror Decision Making: The Story Model, 13 CARDOZO LAW REVIEW 519 (1991).

¹⁸⁷ See, e.g., Smith, How Jurors Make Decisions: the Value of Trial Innovations, in G.T. Munsterman et al, eds., Jury Trial Innovations (1997); B. Michael Dann, “Learning Lessons” and “Speaking Rights”: Creating Educated and Democratic Juries, 68 Indiana Law Journal 1229 (1993);

¹⁸⁸

testimony and defense rebuttal that occur in real trials. Contrast *Punitive Damages*' one-shot-draw-policy-conclusions model of research with the series of experiments by Saks and his co-authors and by Greene and Bornstein, as I discuss in my critique of Chapter 4.¹⁸⁹ These other researchers did look at the effects of defense counter proposals and the effects of defense experts and arrived at considerably different conclusions from those in *Punitive Damages*.

Consistent with MacCoun's view most researchers are careful in generalizing from experimental simulations like those in *Punitive Damages*. Greene and Bornstein's important book, *Determining Damages*, reviewed simulation research on jury decision making on damages, including one of the studies of the Exxon group. In a chapter devoted to reforming damage award decision-making, they stated: "...because we simply do not yet know enough, we urge policy makers and court reformers to tread lightly over this ground that seems ever-shifting and decidedly unstable."¹⁹⁰

As if to put a punctuation point on the Greene and Bornstein's conclusion Professor Jennifer Robbennolt reviewed the overall corpus of experimental research on juries and punitive damages in a recent article to reach the following conclusion:

The research examining the processes by which jurors determine punitive damages suggests that jurors take into account important characteristics of the cases in making their punitive awards.... [J]urors do not appear to make decisions that clearly differ from the decisions that judges would make, certainly not to the extent that most critics of the jury would suggest."¹⁹¹

V. Other Perspectives Bearing on Overreaching Policy Claims in *Punitive Damages*

I have already drawn attention to the fact that Professors Sharkey and Feigenson, after reviewing *Punitive Damages*, separately concluded that its authors have gone beyond their data in making policy recommendations about juries and punitive damages. Professor Feigenson, for example, stated, "...at many points in [*Punitive Damages*'] presentation the data do not support [the authors'] critical view of punitive damages as strongly as they would have readers believe."¹⁹²

In an extremely thorough review of one of the experiments that preceded the book Professor Richard Lempert concluded:

¹⁸⁹ Supra at text around notes ____

¹⁹⁰ Edie Greene and Brian Bornstein, DETERMINING DAMAGES: THE PSYCHOLOGY OF JURY AWARDS (2003).

¹⁹¹ Jennifer Robbennolt, Determining Punitive Damages: Empirical Insights and Implications for Reform, 50 BUFFALO LAW REVIEW 103, 158 (2002).

¹⁹² Feigenson, supra note 21 at 242.

The authors ...do not make an adequate social science case for change [the authors' recommendation that juries should not decide punitive damages], and their recommendation that this should be done deserves *no* weight in any policy arenas.¹⁹³

In the original article containing the experiment that forms the basis of Chapter 9¹⁹⁴ Professor Viscusi drew attention to “the fundamental irrationality that juries display with respect to punitive damage awards” and suggested that proposals to have judges rather than juries award punitive damages would remedy the problem.¹⁹⁵ Steven Garber assessed the experiment’s conceptual adequacy and the methodological soundness of its simulations, concluding that “[t]he supporting argument and evidence are far from compelling,” and that its author’s conclusion about juries was “a key policy recommendation that lacks foundation.”¹⁹⁶ Responding to the same article Robert MacCoun wrote, “Viscusi’s data on their own are clearly too modest to support his sweeping call to either remove punitive damages judgments from the jury or eliminate punitive damages altogether.”¹⁹⁷

VI. Conclusion

Judges and other legal policy makers need to avoid “grasshopper” tendencies in responding to the corpus of work reported in *Punitive Damages*. Various critics of the research have drawn attention to its conceptual and methodological weaknesses and the inappropriateness of the broad policy recommendations set forth by the authors of those experiments. This article has drawn

¹⁹³ Richard Lempert, supra note 88 at 870.

¹⁹⁴ Viscusi, *Corporate Risk Analysis: A Reckless Act?* 52 STANFORD LAW REVIEW 547 (2000).

¹⁹⁵ Id. at 589. In fairness, Professor Viscusi has responded to an earlier draft of the present article by saying that he has never advocated the general abolition of punitive damages. In a personal email communication (March 17, 2004, on file with the author) he drew my attention to the fact that he qualified his view in the Stanford article by referencing his earlier writings advocating only that punitive damages should be abolished for corporate safety and environmental torts. Professor Viscusi pointed out that on page 589 of the article he stated that “Elsewhere, I have proposed abolishing punitive damages for corporate safety and environmental torts,” and referred to his earlier articles in the Georgetown Law Journal. In the email he clarified his position by stating,

“My proposal there is to eliminate punitive damages for specific matters covered by existing federal health, safety, and environmental regulations. Thus, if auto airbags met government standards, then plaintiffs could not obtain punitive damages for injuries caused by airbags. But the fact that NHTSA exists to regulate auto safety does not mean that every auto safety problem is free of possible punitive damages since not every design issue is addressed by specific regulations.”

¹⁹⁶ Steven Garber, *Punitive Damages and Efficiency-promoting Analysis: A Problem with a Solution?* 52 STANFORD LAW REVIEW 1809, 1810 (2000).

¹⁹⁷ Robert MacCoun, *The Costs and Benefits of Letting Juries Punish Corporations: Comment on Viscusi*, 52 STANFORD LAW REVIEW 1821, 1827 (2000).

attention to the failure of the authors of *Punitive Damages* to confront real world data that is inconsistent with their claims. Then, by focusing on the concepts of internal, external and ecological validity it has confronted the simulation experiments on their own terms.

Deconstruction of the methodologies of the experiments demonstrates that they have serious and incontrovertible failings, including limitations of the experimental simulations in comparison to real world juries and biases built into the research materials that did not give a fair test of how jurors might respond to evidence. Considered in conjunction with critiques of *Punitive Damages* or individual experiments in the corpus of Exxon-funded research that have been set forth by other authors, it is abundantly clear that *Punitive Damages* should not be treated as empirical authority for individual cases or for tort reform generally.