Law and Human Behavior, Vol. 23, No. 6, 1999

Epistemological Dilemmas in the Assessment of Legal Decision Making

Robert J. MacCoun[1,2]

Neil Vidmar and Reid Hastie would surely be at the top of anyone's list of the leading scholars in the jury research community. David Schkade and John Payne are major figures in the broader judgment and decision-making community. All four are careful, sophisticated theorists and researchers. Any dispute involving the four of them is bound to be a productive one for the field. I will argue that their dispute brings into focus some fundamental and unresolved epistemological issues at the foundation of the study of legal decision making.

In the abstract to their analyses of verdicts by 121 mock civil juries, Hastie, Schkade, and Payne (1998; henceforth HSP) list their first conclusion: "Most of the mock juries decided that the consideration of punitive damages was warranted, although appellate and trial judges had concluded that they were not warranted" (p. 287). The heart of Vidmar's critique of this claim appears in his title, "Juries Don't Make Legal Decisions!" (exclamation point in the original). Vidmar argues that there is "a fatal conceptual flaw in the experiment that renders it irrelevant to legal policy. Specifically, the experiment asked the jurors to make decisions about law, decisions that are in the exclusive province of the judge. Additionally, the authors created the impression that trial judges, as opposed to juries, were in agreement on the appropriateness of damages and would have 'got it right' if they had decided the cases. In point of fact that was not true in two of the four cases" (p. 705).

In this commentary, I will explain why I think Vidmar's core argument has merit. HSP's attempt to compare mock jury verdicts to an external standard (also see Hastie, Penrod, & Pennington, 1983) is unconvincing because the comparisons are inherently ambiguous. But I will argue that for the field of legal decision making, that is cause for regret and concern—as analysts, we lose a great deal when we fail to find defensible external benchmarks. Moreover, I worry that Vidmar's essay downplays some more convincing evidence that HSP cite to bolster their arguments.

[1]Boalt Hall School of Law and Goldman School of Public Policy, University of California, Berkeley, California.

[2]Correspondence should be addressed to Robert MacCoun, Jurisprudence and Social Policy Program, University of California at Berkeley, 2240 Piedmont Avenue, Berkeley, California 94720; e-mail: maccoun@socrates.berkeley.edu.

0147-7307/99/1200-0723$16.00/1

Before turning to those arguments, it should be noted that, irrespective of the issues Vidmar raises, the HSP paper makes many useful empirical contributions. First, it extends the growing literature on individual difference correlates of civil juror judgments, replicating the general finding that easily measured sociodemographic variables offer very limited predictive power. It replicates earlier findings on the correlates of foreperson choice, and extends that literature in an important way with some novel analyses of foreperson influence (or lack thereof). It provides important new evidence on deliberation dynamics and judgmental shifts in civil cases. Finally, it contributes valuable evidence on juror legal comprehension and on deliberation content. Of special interest is the use HSP make of the latter results to interpret jury verdicts—a point I return to below.

THE PRAGMATICS OF JURY REASONING

My own reading is that the line separating the roles of jury and judge—in practice, if not in legal tradition—is blurrier than Vidmar implies. Nevertheless, I shared his reaction to the HSP comparison of mock jury verdicts and appellate rulings in the stimulus cases. Indeed, I would offer another line of argument that bolsters his objection. Even without knowing the details Vidmar provides about dissenting appellate opinions—and in fairness, HSP do say "we have established a consistent discrepancy between jurors' and *some* judges' judgments for these cases" (p. 306, my emphasis)—I think there are other grounds for skepticism.

For example, recent work on pragmatic conversational norms (Hilton, 1995; Schwarz, 1996) suggests that many (but not all) experimental demonstrations of bias or error in the literatures on attribution and on judgment under uncertainty are vulnerable to a rival interpretation. Specifically, it is conceivable that experimental participants "read between the lines" and infer, from the experimental situation, premises that are not actually given in the written stimulus materials. If so, their judgments may reflect a coherent, pragmatic logic not captured by formal normative analyses of the decision tasks. In the HSP experiment, mock jurors are asked to judge liability for punitive damages, and the magnitude of any such damages, by the experimenters and (hypothetically) by the trial judge in each stimulus case. From a pragmatic reasoning perspective, it seems likely that their mock jurors might have concluded that the cases at hand already passed a legal threshold of eligibility for punitive damages—conditional of course on the jury's own judgments of malice or recklessness as requested in the judge's charge. Otherwise, why would the (simulated) judge and (real) experimenter be asking? I have no evidence that this is what in fact occurred in their experiment; I simply offer it as a plausible rival interpretation of their liability rates.

An added concern is that the HSP stimulus materials, though based closely on actual trials, were nevertheless simulations of those trials. It is conceiveable that the cases were subtly (and unintentionally altered) in significant ways as a result of the necessary abstraction involved in reducing the trial materials to a manageable, standardized format for the experiment. In general, mock jury experiments are ill-suited to estimating the magnitude of actual liability rates or other effect sizes

(MacCoun, 1993; also see Cooper & Richardson, 1986). Granted, HSP might respond that the abstraction process *advantaged* mock jurors relative to actual trial jurors, but we cannot know that for sure.

THE COHERENCE AND CORRESPONDENCE THEORIES OF TRUTH

These concerns strike me as serious challenges to the HSP conclusions, but I think the HSP study draws our attention to deeper, more fundamental issues in psycholegal research. To understand those issues, it is useful to draw upon the traditional philosophical distinction between the *coherence* and *correspondence* "theories of truth." I put the latter phrase in quotes because I see these as methodological strategies (Hastie & Rasinski, 1988) rather than theories of truth—i.e., they seem more epistemological than ontological. According to Hammond (1996, p. 106), "Correspondence theory focuses on the *empirical* accuracy of judgments, irrespective of whether the cognitive activity of the judge can be justified or even described. . . . Coherence theorists have opposite interests; they examine the question of whether an individual's judgment processes meet the test of rationality—internal consistency—irrespective of whether the judgment is empirically accurate."

Hastie and Rasinski (1988) have expanded on this distinction in a taxonomy of "logics" for assessing accuracy in judgment research, one I have found quite useful in my own research (Kerr, MacCoun, & Kramer, 1996; MacCoun, 1998). Two of these logics illustrate the coherence approach, as it is commonly practiced in attribution research, the "heuristics and biases" tradition, and almost all mock jury research (see the review in Kerr et al., 1996). In these two logics, the presence and content of various informational cues is manipulated in a between- or within-subjects experiment. In the first version, a bias is established by showing that a judge is "using a bad cue"—i.e., overutilizing a cue relative to normative standards (e.g., legal rules of evidence, a rational choice model, or the cue's objective predictive validity). In the second, a bias is established by demonstrating that the judge is "missing a good cue"—underutilizing a cue relative to normative standards. These are "sins of commission" and "sins of omission," respectively (Kerr et al., 1996).

A third logic exemplifies the correspondence approach: bias or error is established directly by measuring the discrepancy between the judgment and the true state being judged. This logic has been quite fruitful in psychophysics, but perhaps less so in social psychology, where we often lack objective measures of the "true" state of the sociopolitical environment. Though they did not say so explicitly, I believe this is what HSP were striving for in their assessment of liability rates. And there are good reasons why we should want a correspondence analysis. The issue here is not external validity in the generalizability sense, which in my opinion is an overrated concern in mock jury research (Bornstein, 1999; MacCoun, 1993).[3]

[3]I do not think this contradicts my earlier claim that simulated cases cannot provide good effect size estimates. Mock jury experiments are best used to test hypotheses, which generally pertain to the existence and direction of effects rather than their magnitude. Arguably, econometric analyses of actual jury awards are better suited to estimating magnitudes of effects, or perhaps meta-analyses across a sufficiently large and heterogeneous sample of mock jury experiments.

Instead, the question is whether we can ground our evaluations in some objective standards for accuracy.

Dawes (1993, p. 227) has called on jury researchers to pay more attention to the need for correspondence tests:

> I would like to suggest that deliberately choosing cases that are 'ambiguous' may not be the best way of studying judgment, particularly not in the jury decision context ... So why not use trials where there is a criterion of truth or correctness? Certainly there are cases in which learned jurors would agree that one verdict is 'right' and another is 'wrong.' ... In general ... investigation of human judgment on tasks with a criterion (e.g., an actual disease, actual success or failure at staying out of jail or in graduate school) has led to greater insight about the judgmental process than have investigations of tasks without one (e.g., aesthetic preference).

Reading Hammond and Dawes, I am fully persuaded of the advantages offered by a correspondence approach. Unfortunately, I see three enormous obstacles to Dawes' recommendation. First, many reliable judgmental phenomena in the jury literature are most likely to be observed in ambiguous cases and will be overlooked in clear cases; see Kerr et al. (1996) for a review and a formal mathematical analysis. Second, though they overstated their position, Priest and Klein (1984) were surely correct to argue that the kinds of cases that come to jury trial are unlikely to be a random sample of all cases; because the lopsided cases will tend to be resolved by pretrial settlement (civil) or plea bargaining (criminal), jury trials will tend to overrepresent the most ambiguous, difficult cases. Finally, the third problem is the most fundamental one: If we had an external criterion for judging correct verdicts, why wouldn't we just use it to resolve the case? Why have a jury trial in the first place?

THE INHERENT DIFFICULTIES IN COMPARING JUDGE AND JURY JUDGMENTS

That brings me to the fourth and most problematic Hastie–Rasinski "logic": A direct comparisons of judgments across decision makers (or across groups of decision makers that differ in some attribute; e.g., men vs. women, liberals vs. conservatives, etc.). With this approach, if the judgments are discrepant, then even in the absence of external criteria, one can arguably infer bias. (Bayesians would disagree, provided that the difference lies in prior probabilities rather than judgments of the diagnosticity of the evidence.) The problem is, bias on who's part? A weakness of this logic is that the observed discrepancy tells us nothing about whether either decision maker (or group) is actually accurate; both could be wrong (MacCoun, 1998).

It seems to me that HSP confound the first and fourth logics. It would appear that, following Dawes and Hammond, they want a correspondence test—a direct comparison of the jury verdicts and "the true state" of the world. But their comparison ultimately suffers from all the problems of the fourth logic. And in this case, there is a deeper normative problem. They are comparing two populations—lay judgments and expert judicial judgments—in a domain in which each of those populations is normatively advantaged in different ways (MacCoun, 1993). Judges

are clearly more expert in the law, and in fact HSP demonstrate a staggering lack of comprehension of that law by their lay jurors. But jurors are better suited for reflecting community standards in an ambiguous domain like tort litigation.

In a detailed examination of judge and jury roles in punitive damage cases, Scheiner (1991) notes that "jury bias can mean that the jury exercises sound discretion, but that jury determinations nevertheless systematically differ from those that judges would make because of moral or experiential differences between judges and juries. Because there is no unbiased baseline, any claim of juror bias of this ... type can be restated as an equally true claim of judicial bias" (p. 165). "Fig. 1 depicts four different inferences one might draw from judge-jury disagreement. Depending on its source, such disagreements might reflect jury incompetence, jury lawlessness or nullification, unintended bias, or legally acceptable differences in values (see Kalven & Zeisel, 1966; Scheiner, 1991). This is not to say that judge–jury comparisons are not, or should not be, of enormous interest to jury researchers (see reviews by, e.g., Kalven & Zeisel, 1966; MacCoun, 1993; Scheiner, 1991). But the discrepancies do not speak for themselves. Taken by itself, the fact that a majority of the HSP juries favored punitive damages in three of four cases where judicial review ruled otherwise is not on its face a clear indictment of the jury system.

For me, this concern first arose in Hastie's landmark 1983 book with Steve Penrod and Nancy Pennington. Noting that only 39 of 69 mock juries reached a second-degree murder verdict in the armed robbery case used in their study, they argued that

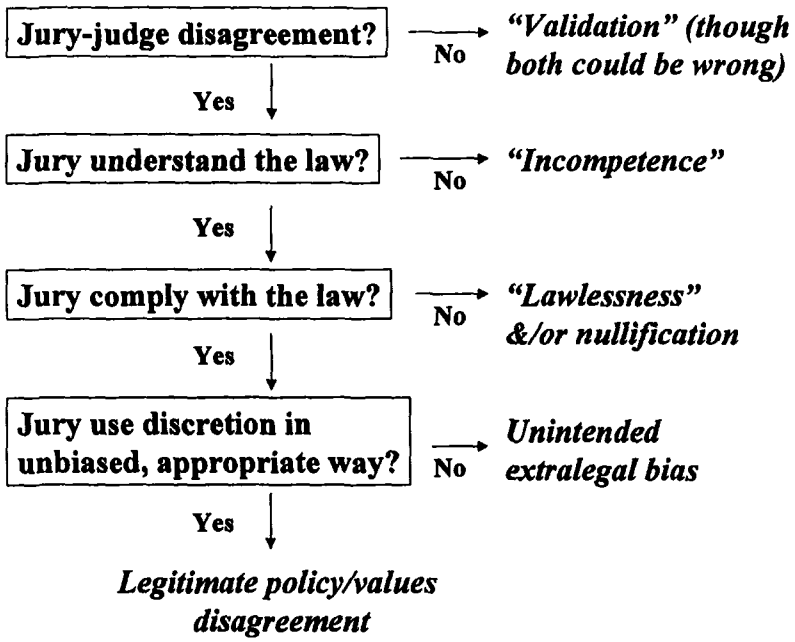

Fig. 1. Possible sources of judge–jury disagreement.

> Deviations from the second-degree murder verdict for the stimulus case are a cause for concern, because this verdict is the 'correct' one on a number of grounds. First, the jury in the actual trial on which the case was based returned a verdict of second degree murder. Second, the legal experts who viewed the case consistently selected the second-degree murder verdict. Third, second degree murder is the modal verdict for the experimental juries (Hastie, Penrod, & Pennington, 1983, p. 62).

Given the variance in verdicts among their 69 mock juries, it seems strange to grant the actual jury's verdict any particular normative standing. Their other two arguments are also problematic. If jury decision making is what is being assessed, why should the "modal verdict" be an indicator of truth? If there is a systematic flaw in jury judgment, it might well be reflected in the modal verdict. And again, if legal experts can determine the correct verdict, why did we need a jury trial?

But importantly, in their next sentence Hastie et al. (1983) backed away from a purely correspondence-based interpretation and shift to a coherence-based interpretation: "However, the issue of the correctness of the final verdict cannot be resolved in absolute terms, for there are no ideal rational or empirical criteria for accuracy in jury decisions. At a minimum, each jury's evaluations of witness credibility, reasonable inferences from the testimony, and reconstruction and interpretation of the judge's summary of the law must be considered in order thoroughly to assess verdict correctness" (p. 62).

I think the totality of the HSP paper can be read as offering the same argument, and here is where I think Vidmar sells their recent study short. HSP state that "our primary goal is to sketch an initial picture of the juror and jury cognitive and social processes when making a typical serious civil decision" (p. 289), and their abstract states that "the tendency to find the defendant liable was partly due to jurors' failure systematically to consider the full set of legally necessary conditions for the verdicts they rendered" (p. 287). They document dismally low legal comprehension, and demonstrate a striking neglect of legally significant concepts (maliciousness, recklessness/callous disregard) during deliberation. These facts cannot unequivocally establish jury incompetence in civil cases—and granted, the low comprehension may be partly the fault of the judge's instructions and the state of tort law—but these findings seem too troubling to dismiss.

But holding judicial judgment (roughly) constant while analyzing within- and between-jury variance is a weak basis for comparing judge and jury judgment. A more convincing approach would be a variant on Kalven and Zeisel (1966)—a comparison of judge and jury judgments across a large sample of cases, with statistical analysis of the correlates of disagreement. Better yet is an experimental approach, in which judges and jurors (or juries) are asked to respond to the same stimulus materials, the same manipulations, and the same process-oriented dependent measures (e.g., Howe & Loftus, 1992; Wells, 1992)—an approach that Hastie has used in other recent work (Hastie & Viscusi, 1998).

THE POLITICS OF CIVIL JURY RESEARCH

Vidmar mentions in passing (footnote 11) the fact that the Exxon Corporation sponsored the HSP study. The question of biased evidence interpretation is complex;

I address it at length elsewhere (MacCoun, 1998). Here I simply agree that the issue of research funding is a serious one for the psychology-law community. Yet sociolegal scholars without such funding often seem blithely unconcerned with their own partisan tendencies. To the extent that there's bias in the jury research community as a whole, the net slant is arguably pro-rather than anti-jury. I suspect many of us are still reacting against the shrill, distorted, and omnipresent tort reform media campaign of the 1980s. But for researchers, both advocacy and counteradvocacy are normatively unjustifiable unless one's advocacy stance is made explicit (MacCoun, 1998). If perfect objectivity is impossible, our greatest fear should be a premature homogeneity of viewpoints and arguments. Given the inherent difficulty in anchoring jury research to objective standards of truth, some disagreements among researchers strikes me as a healthy development.

REFERENCES

Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior, 23,* 75–92.

Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 71,* 179–184.

Dawes, R. M. (1993). Notes on the sampling of stimulus cases and the measurement of responses in research on juror decision making. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 225–228). New York: Cambridge University Press.

Greene, E., & Loftus, E. F. (1998). Psycholegal research on jury damage awards. *Current Directions in Psychological Science, 7,* 50–54.

Hammond, K. R. (1996). *Human judgement and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* New York: Oxford University Press.

Hammond, K. R., Harvey, L. O., & Hastie, R. (1992). Making better use of scientific knowledge: Separating truth from justice. *Psychological Science, 3*(2), 80–87.

Hastie, R., Penrod, S. D., & Pennington, N. (1983). *Inside the jury.* Cambridge, MA: Harvard University Press.

Hastie, R., & Rasinski, K. A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 193–208). Cambridge: Cambridge University Press.

Hastie, R., Schkade, D., & Payne, J. (1998). A study of juror and jury judgments in civil cases: Deciding liability for punitive damages. *Law and Human Behavior, 22,* 287–314.

Hastie, R., & Viscusi, K. (1998). What juries can't do well: The jury's role as risk manager. *Arizona Law Review, 40,* 901–921.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118,* 248–271.

Howe, E. S., & Loftus, T. C. (1992). Integration of intention and outcome information by students and circuit court judges: Design economy and individual differences. *Journal of Applied Social Psychology, 22,* 102–116.

Kalven, H., & Zeisel, H. (1966). *The American jury.* Boston: Little, Brown.

Kerr, N., MacCoun, R. J., & Kramer, G. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review, 103,* 687–719.

MacCoun, R. J. (1989). Experimental research on jury decision making. *Science, 244,* 1046–1050.

MacCoun, R. J. (1993). Inside the black box: What empirical research tells us about decisionmaking by civil juries. In R. E. Litan (Ed.), *Verdict: Assessing the civil jury system* (pp. 137–180), Washington, DC: The Brookings Institution.

MacCoun, R. (1998). Biases in the interpretation and use of research results, *Annual Review of Psychology, 49,* 259–287.

Priest, G. L., & Klein, B. (1984). The selection of disputes for litigation. *Journal of Legal Studies, 13,* 1–55.

Scheiner, A. H. (1991). Judicial assessment of punitive damages, the Seventh Amendment, and the politics of jury power. *Columbia Law Review, 91,* 142–226.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation.* Mahwah, NJ: Erlbaum.
Vidmar, N. (1999). Juries don't make legal decisions! And other problems: A critique of Hastie et al. on punitive damages. *Law and Human Behavior,* this issue.
Vidmar, N., & Rice, J. L. (1993). Assessments of noneconomic damage awards in medical negligence—A comparison of jurors with legal professionals. *Iowa Law Review, 78,* 883–911.
Wells, G. L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62,* 739–752.