COMPARING LEGAL FACTFINDERS: REAL AND MOCK, AMATEUR AND PROFESSIONAL

ROBERT J. MACCOUN*

The excellent articles by Brian Bornstein and Sean McCabe[1] and Jennifer Robbennolt[2] in this Symposium document the current state of two longstanding debates in jury scholarship—one on the external validity of mock jury experiments, the other on the relative fact-finding ability of juries versus judges. Over a decade ago, I examined the state of evidence in each debate.[3] With respect to what Bornstein and McCabe refer to as "consequentiality"—the question of whether decisions result in real outcomes[4]—in 1989 I noted that "[e]xperiments comparing mock jurors with subjects who thought they were actually trying a case have been inconclusive; different studies have found mock jurors' verdicts to be more lenient, less lenient, and no different from those of 'actual' jurors."[5] Bornstein and McCabe note with some frustration that there is no new evidence to clarify these results[6]—indeed, no studies have examined the issue since 1986.[7] In stark contrast, Robbennolt shows that there has been considerable empirical progress on the second issue since 1993,[8] when I was able to identify only two direct experimental comparisons of judge and juror judgmental performance.[9]

Perhaps framing these questions as "debates" is misleading, since in neither case can we expect empirical research to provide an unequivocal, dichotomous verdict like "mock jury research is valid" or "judges are superior." Both issues involve inevitable trade-offs among goals—epistemological goals in the first case and legal policy goals in the second.

I. THE REALISM OF JURY RESEARCH

The "realism" of the mock jury paradigm involves at least three dimensions: whether the "stimulus case" materials resemble actual trials, whether the samples resemble real jury-pool populations, and whether the psychology of assessing fictional cases resembles what jurors experience when reaching real verdicts involving real litigants. Brian Bornstein has already made a major contribution to our understanding of the first two dimensions; his 1999 meta-analysis of twenty years of mock jury research found little indication that either stimulus case realism (for example, paper-pencil versus audiotape versus videotape) or study population (for example, student versus

* Professor of Law and Professor of Public Policy, University of California at Berkeley.

1. Brian H. Bornstein & Sean G. McCabe, *Jurors of the Absurd? The Role of Consequentiality in Jury Simulation Research*, 32 FLA. ST. U. L. REV. ___ (2005).

2. Jennifer K. Robbennolt, *Evaluating Juries by Comparison to Judges: A Benchmark for Judging?*, 32 FLA. ST. U. L. REV. ___ (2005). In the interest of full disclosure, I recently coauthored a paper with Robbennolt. *See* Jennifer K. Robbennolt, John M. Darley & Robert J. MacCoun, *Symbolism and Incommensurability in Civil Sanctioning: Decision Makers as Goal Managers*, 68 BROOK. L. REV. 1121 (2003).

3. Robert MacCoun, *Inside the Black Box: What Empirical Research Tells Us About Decisionmaking by Civil Juries*, *in* VERDICT: ASSESSING THE CIVIL JURY SYSTEM 137 *passim* (Robert E. Litan ed., 1993) [hereinafter MacCoun, *Inside the Black Box*]; *see also* Robert J. MacCoun, *Experimental Research on Jury Decision-Making,* 244 SCIENCE 1046, 1046-50 (1989) [hereinafter MacCoun, *Experimental Research*].

4. *See* Bornstein & McCabe, *supra* note 1, at footnote 35-38.

5. MacCoun, *Experimental Research*, *supra* note 3, at 1046.

6. Bornstein & McCabe, *supra* note 1, at footnote 39-43.

7. *Id.* at around footnote 43 tbl. 1.

8. Robbennolt, *supra* note 2, at ____.

9. *See* MacCoun, *Inside the Black Box*, *supra* note 3, at 166 (citing Edmund S. Howe & Thomas C. Loftus, *Integration of Intention and Outcome Information by Students and Circuit Court Judges: Design Economy and Individual Differences*, 22 J. APPLIED SOC. PSYCHOL. 102 (1992); and Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. PERSONALITY & SOC. PSYCHOL. 739 (1992)).

general community) systematically influences research conclusions.[10] It is difficult to avoid the conclusion that efforts to maximize realism on these dimensions have more to do with research marketing than scientific validity; whether the marketing payoff is worth the tremendous opportunity costs is an open question that funders ought to take seriously.[11]

Readers outside the jury research community might be surprised to find so little attention given to the third dimension—the consequentiality issue. But this neglect is understandable, given the small size and limited resources of the jury research community. A handful of prior studies with inconsistent results provides little incentive for investing the considerable effort in a new experiment, unless something important is at stake—an unresolved legal policy debate (for example, jury size or bifurcating damages) or a competition among basic theories of juror judgment (for example, "cognitive averaging" versus story-based or constraint satisfaction models[12]). Perhaps new research on consequentiality would persuade policymakers to take mock jury research more seriously—at least if consequentiality proved inconsequential. But the issue seems unlikely to be the only (or even a major) obstacle to having our research findings drive major legal policy decisions. Far from being unique to the mock jury literature, the complaint that policymakers ignore social science research seems nearly universal.[13]

A lack of real decision consequences, on its face, seems relevant to the assessment of mock jury research. But as Bornstein and McCabe point out, there are qualitatively distinct ways in which outcomes in mock and real jury trials might systematically differ: a variable might have a greater effect on verdicts in one setting than in the other, it might influence outcomes in one setting but have no effect on other, or it might affect each setting in opposite ways.[14] Though it is easy to see how realism might influence the magnitude of particular correlations among evidentiary or procedural factors and verdicts,[15] it is hard to identify compelling reasons why mock and actual juror reasoning might differ systematically in kind, rather than degree. Various theories in the behavioral science judgment and decision literatures address the realism question, but as far as I can tell, none predict important qualitative differences of the sort that would make mock jury research misleading.

The most straightforward possibility is that real jurors may try harder than mock jurors—think harder, deliberate longer, and ponder more deeply. The widely influential "elaboration likelihood model" of Richard Petty and John Cacioppo has stimulated a great deal of research on the effects of motivation on reasoning and persuasion.[16] That work shows that information in the environment (whether verbal arguments, visual evidence, or other nonverbal stimuli) can influence opinion formation via one of two qualitatively different routes.[17] The "central route" involves systematic

---

10. Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?*, 23 LAW & HUM. BEHAV. 75 (1999).

11. Videotaped trial reenactments, the use of jury pool volunteers, and mock jury deliberations conducted in courthouse settings add greatly to the expense and duration of an experiment; a junior researcher could plausibly conduct three or four low-fidelity experiments in university settings with the same or less time and money.

12. *See* INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING (Reid Hastie ed., 1993) [hereinafter INSIDE THE JUROR]; MacCoun, *Experimental Research*, *supra* note 3; Robbennolt, Darley & MacCoun, *supra* note 2.

13. *See, e.g.*, DAVID GREENBERG ET AL., SOCIAL EXPERIMENTATION AND PUBLIC POLICYMAKING (2003); RICHARD P. NATHAN, SOCIAL SCIENCE IN GOVERNMENT: USES AND MISUSES (1988); THE POLITICS OF NUMBERS (William Alonso & Paul Starr eds., 1987); Robert J. MacCoun, *Biases in the Interpretation and Use of Research Results*, 49 ANN. REV. PSYCHOL. 259 *passim* (1998).

14. Bornstein & McCabe, *supra* note 1, at ____; *see also* Robert M. Bray & Norbert L. Kerr, *Use of the Simulation Method in the Study of Jury Behavior*, 3 LAW & HUM. BEHAV. 107 (1979).

15. *E.g.*, Bornstein, *supra* note 10; Nancy Mehrkens Steblay et al., *The Effects of Pretrial Publicity on Juror Verdicts: A Meta-Analytic Review,* 23 LAW & HUM. BEHAV. 219 (1999).

16. *See* Richard E. Petty & John T. Cacioppo, *The Elaboration Likelihood Model of Persuasion*, *in* 19 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 123 (Leonard Berkowitz ed., 1986). For a very similar theory and an extensive literature review, see ALICE H. EAGLY & SHELLY CHAIKEN, THE PSYCHOLOGY OF ATTITUDES 305-25 (1993).

17. Indeed, perhaps the most important development in psychology in the past decade or so has been the convergent discovery (across various research domains) that the brain has at least two qualitatively different modes of reasoning—one fast, automatic, associative, and largely unconscious; the other slow, controlled, verbal, and largely conscious. *See* KEITH E. STANOVICH, THE ROBOT'S REBELLION: FINDING MEANING IN THE AGE OF DARWIN 31-79 (2004).

"elaboration" of information—that is, "thinking." The "peripheral route" involves a casual, and usually temporary, coloring of one's views due to superficial cues (such as physical attractiveness or humor) or heuristic "rules of thumb" (such as "more arguments are better than fewer arguments"). Peripheral cues are most influential when the persuasion target's motivation or ability to scrutinize logical arguments is low. Argument quality becomes more important when motivation and ability are high. But studies testing this theory mostly examine reactions to consumer advertising, rather than explicit decision problems involving intellectual or moral judgment. The mock jury task, by contrast, is highly engaging. In recent mock jury research, even participants with relatively low motivation to scrutinize evidence were sensitive to argument quality.[18] And as Bornstein and McCabe note, considerable research on the effects of incentives for accuracy—even extremely large monetary rewards——has failed to show any consistent effect on decision quality in cognitively engaging judgment tasks.[19]

Another possibility is that motivation favors a particular verdict option. For example, based on the standard decision theoretic account of trade-offs among false-positive and false-negative errors,[20] Bornstein and McCabe argue that if jurors are assumed to focus on avoiding false convictions of innocent people, then the greater the penalty severity, the greater the perceived cost of error.[21] As this perceived cost of an error increases, jurors should require more evidence of guilt before voting to convict. With this shift in criterion, jurors should therefore be less likely to vote guilty when the penalty is more severe.[22]

Ceteris paribus, this is a straightforward decision theory prediction. But what it ignores is the fact that the other decision error—falsely acquitting a guilty defendant—is also considerably more costly in actual trials. Thus, decision theory is mute unless there are compelling reasons to believe that the effects of realism are asymmetric, increasing the cost of false positives more than the cost of false negatives. I suppose it is an empirical question, but a priori I see no particular reason to expect such an asymmetric effect.

Another relevant theoretical account is Irving Janis and Leon Mann's 1977 conflict model.[23] A full presentation is beyond the scope of this Comment, but in a nutshell, Janis and Mann distinguish three styles of decision under conflict and stress: *vigilance* (a careful and systematic weighting of costs and benefits), *hypervigilance* (roughly, "panic"), and *defensive avoidance* (preemptive efforts to rationalize or bolster what is expected to be an inevitable bad outcome).[24] Janis and Mann argue that vigilance can only occur under conditions of moderate stress and when there is an adequate time period for gathering and weighing information.[25] When there is insufficient time, they predict hypervigilance.[26] When every choice poses serious risks with no apparent solution, they predict defensive avoidance.[27] The Janis and Mann model has served an important role in helping researchers understand how real-world decision processes may depart from the standardized and idealized conditions of our laboratory experiments. Yet the model has no obvious implications for the comparison of mock and real jury trials. Real jury trials occur at a stately, almost glacial pace, without the urgency and immediacy faced by decisionmakers in the case studies examined by Janis and Mann. And while it may be that no verdict option offers a clear guarantee of suc-

---

18. Donna Shestowsky & Leonard M. Horowitz, *How the Need for Cognition Scale Predicts Behavior in Mock Jury Deliberations*, 28 LAW & HUM. BEHAV. 305, 333 (2004).

19. Bornstein & McCabe, *supra* note 1, at pgh of footnote 11; *see also* Colin F. Camerer & Robin M. Hogarth, *The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework*, 19 J. RISK & UNCERTAINTY 7 (1999).

20. *See* INSIDE THE JUROR, *supra* note 12; MacCoun, *Experimental Research*, *supra* note 3.

21. Bornstein & McCabe, *supra* note 1, at IIIB.

22. *Id.*

23. IRVING L. JANIS & LEON MANN, *A Conflict Model of Decision Making*, *in* DECISION MAKING: A PSYCHOLOGICAL ANALYSIS OF CONFLICT, CHOICE, AND COMMITMENT 45, 45-80 (1977).

24. *Id.* at 52-64.

25. *Id.* at 59.

26. *Id.*

27. *Id.* at 57-58.

cess, as noted above, there is no clear reason why this is any more or less true in real trials than in the mock jury situation. Jennifer Lerner and Philip Tetlock have offered a related account of the effects of real-world pressures on human decisionmaking, focusing on the role of accountability.[28] They argue that the effects of expecting to be held accountable for a decision will vary depending on whether the 'decisionmaker knows the views of the various stakeholders and other audiences that will scrutinize the decision.[29] If the audience views are unknown, the decisionmaker will be motivated to engage in "preemptive self-criticism"[30]—analogous to the "vigilant" style of careful reasoning described by Janis and Mann.[31] If the audience views are known, the decisionmaker will engage in some "attitude shifting"—moving in the direction of the audience's viewpoint.[32] Presumably, real jurors are more accountable for their decisions than mock jurors; the stakes are higher and trials, by their very nature, are highly public acts.[33] But the filtering process that produces trials almost guarantees that the "audience" will not share a single point of view. Litigants are under enormous pressure to settle or plea bargain when the evidence is stacked against them, and, to some extent, changes of venue and venire compensate for overwhelming public prejudice at odds with actual evidence strength. Again, it is hard to find clear a priori grounds for rejecting mock jury data on the basis of accountability effects.

In sum, there are a variety of theoretical perspectives in the judgment literature that suggest ways in which real-world behavior might depart from stylized laboratory tasks. But none of those theories offer clear predictions about the effects of jury verdict consequentiality. That is not a shortcoming of the theories; rather, the variables they highlight are not variables that cleanly differentiate mock juries from real juries.

Implicit in the debate about mock jury realism is the notion that there are better research alternatives. The mock jury method is sometimes viewed as a "second-best" response to legal and ethical barriers to observing actual jury deliberations. But in point of fact, even absent those barriers, actual jury trials have many limitations for scientific research. There are inherent trade-offs among behavioral science methods. Statistical analyses of actual jury verdicts, post-trial juror interviews, shadow juries, field experiments, and mock jury experiments each offer certain advantages, but also certain weaknesses.[34] For example, statistical analysis of actual trial outcomes is an important research tool but is notoriously weak at establishing causation.[35] Experimentation—specifically, random assignment to condition—is a powerful inferential tool for addressing causal questions that purely statistical analyses cannot answer. But there are many logistical, legal, and ethical barriers to the experimental manipulation of variables of interest in actual trials, such as new rules of evidence, new procedures, litigant characteristics, or variations in trial evidence. We should not rely solely on mock jury research, nor should we abandon it in the belief that "more realistic" methods are necessarily less vulnerable to faulty inference. Empirical progress requires triangulation across fallible sources of evidence.

---

28. *See* Jennifer S. Lerner & Philip E. Tetlock, *Accounting for the Effects of Accountability*, 125 PSYCHOL. BULL. 255 (1999).
29. *See id.* at 256-57.
30. *Id.* at 257.
31. *See* JANIS & MANN, *supra* note 23, at 59.
32. Lerner & Tetlock, *supra* note 28, at 256. Tetlock also considered what happens when one only discovers that one is accountable after a decision has already been rendered. In such situations, accountability cannot affect the decision, but if the audience is dissatisfied, the decisionmaker may engage in "defensive bolstering" in an effort to rationalize the position—analogous to Janis and Mann's notion of defensive avoidance. *See id.* at 257-58; JANIS & MANN, *supra* note 23, at 57-58.
33. This is mitigated somewhat by the partial anonymity of jurors and the fact that they are not formally required to explain their verdicts. But increasingly aggressive media coverage is undermining both these caveats.
34. MacCoun, *Inside the Black Box*, *supra* note 3.
35. For example, alleged "deep-pocket" effects in compensatory damage awards apparently result from the real-world confounding of defendant wealth with other factors that distinguish corporations from individual people. *See* Robert J. MacCoun, *Differential Treatment of Corporate Defendants by Juries: An Examination of the "Deep-Pockets" Hypothesis*, 30 LAW & SOC'Y REV. 121 (1996).

II. JUDGES VERSUS JURIES

There is an entirely different sense in which the lack of grounding in "reality" might hinder sociolegal research. In many areas of research on judgment and decisionmaking, progress is made by comparing a decisionmaker's judgments to some independently determined criterion of "truth": Did the radiologist correct identify a cancerous tumor? Did it rain when the meteorologist said it would? This approach relies on epistemology's traditional "correspondence theory" of truth.[36] In sociolegal research, we rarely have any independent criterion—if we did, we probably would not rely on trials, at least not for factual questions.[37] Instead, we either compare types of fact-finders directly, or we compare factfinder performance to some normative theory that prescribes appropriate versus inappropriate sources of information and rules for combining that information.[38]

The recent judge-juror comparisons reviewed by Jennifer Robbennolt are noteworthy for use of both of these latter strategies in combination.[39] The direct judge-jury comparisons tell us whether these actors differ, but such comparisons do not tell us who was performing better; for example, we cannot simply assume that whatever the judges favor is the better judgment, unless perhaps the judgment involves a narrow legal technicality. But the use of experimental tasks grounded in normative theories of inference is allowing researchers to calibrate these judge-jury differences relative to some independent standard for judgment. Of course, these inferential standards address only the inferential dimension of legal factfinding; a full comparison of judges and juries must consider other factors, such as community representation, perceived fairness and legitimacy, cost, and efficiency.[40]

Testing performance relative to explicit theories of performance may offer a rhetorical advantage in addition to any scientific benefits. It may be naïve for mock jury researchers to hope to influence policy using our empirical findings. Empirical research by econometricians plays an important role in legal policy, but its influence surely pales by comparison to the theoretical impact of the law and economics movement, which emphasizes ideas over data. Mock jury experiments are particularly well-suited for testing our ideas about how legal factfinding occurs, but if we want to influence legal policymaking, offering good arguments may be more powerful than defending arguably good data.

---

36. *See* KENNETH R. HAMMOND, HUMAN JUDGMENT AND SOCIAL POLICY: IRREDUCIBLE UNCERTAINTY, INEVITABLE ERROR, UNAVOIDABLE INJUSTICE (1996).

37. *See* Robert J. MacCoun, *Epistemological Dilemmas in the Assessment of Legal Decision Making*, 23 L. & HUM. BEHAV. 723, 728 (1999).

38. For a taxonomy of strategies for assessing judgment accuracy, see Reid Hastie & Kenneth A. Rasinski, *The Concept of Accuracy in Social Judgment*, *in* THE SOCIAL PSYCHOLOGY OF KNOWLEDGE 193 (Daniel Bar-Tal & Arie W. Kruglanski eds., 1988).

39. *See* Robbennolt, *supra* note 2.

40. For public views of how judges and juries differ on these dimensions, see Robert J. MacCoun & Tom R. Tyler, *The Basis of Citizens' Perceptions of the Criminal Jury: Procedural Fairness, Accuracy, and Efficiency*, 12 LAW & HUM. BEHAV. 333 (1988).