

ASKING THE RIGHT QUESTIONS ABOUT JUDGE AND JURY COMPETENCE

GREGORY MITCHELL*

We should ask at least two questions when considering taking a task away from juries and giving it to judges on grounds that juries perform the task suboptimally:¹ Is there good reason to believe that judges will perform the task better? What perverse effects may we introduce with this change in tasks?² Despite the commonsense basis for these questions, surprisingly little research exists to answer either one.³ Fortunately, Jennifer Robbennolt's contribution to this symposium issue provides a very helpful synthesis of the existing empirical research that compares the decisions of judges and juries, which leads to a tentative negative answer to the first question,⁴ and elsewhere Robbennolt partially answers the second question.⁵ Professor Robbennolt asks and answers the right questions about the allocation of duties between jury and judge, but we need much more research on judicial competence to correct the imbalance relative to the large amount of research on jury competence.⁶

* Associate Professor of Law, Florida State University College of Law. This paper benefited from discussions at the symposium on the Behavioral Analysis of Legal Institutions held at the Florida State University College of Law.

1. Most notable in recent times is the considerable effort directed at reducing the jury's role in setting punitive damages because jurors supposedly perform this task poorly. See, e.g., Cass R. Sunstein, *What Should Be Done?*, in PUNITIVE DAMAGES: HOW JURIES DECIDE 242, 242 (2002) ("We have seen that jurors face many problems in trying to generate a sensible system of punitive damage awards."); Lisa M. Sharkey, Comment, *Judge or Jury: Who Should Assess Punitive Damages?*, 64 U. CIN. L. REV. 1089, 1089 (1996) ("Over the past two decades, state legislatures throughout the nation have sharply altered the manner in which punitive damages are awarded. . . . As part of this trend, the state legislatures of several states took from jurors the power to assess punitive damages and placed it in the hands of the trial judge." (footnotes omitted)).

2. Cf. Reid Hastie, *Putting It All Together*, in PUNITIVE DAMAGES: HOW JURIES DECIDE, *supra* note 1, at 211, 234 ("If we are critical of the jury's performance when deciding on punitive damages, it is important to ask: What is the alternative to a jury?").

3. See Jennifer K. Robbennolt, *Evaluating Juries by Comparison to Judges: A Benchmark for Judging?*, 32 FLA. ST. U. L. REV. 469, 470-71 & nn. 5-6 (2005) (summarizing the state of the research on judge versus jury competence).

4. See *id.* at 509 ("While there appear to be some differences in outcomes under some circumstances, judges and jurors generally appear to be influenced by similar factors and suffer from many of the same difficulties in making their decisions."). For a recent study presenting evidence that juries award punitive damages more frequently and in higher amounts than judges, see Joni Hersch & W. Kip Viscusi, *Punitive Damages: How Judges and Juries Perform*, 33 J. LEGAL STUD. 1 (2004).

5. On possible perverse effects from limiting the authority of juries to set punitive damages, see Jennifer K. Robbennolt, *Determining Punitive Damages: Empirical Insights and Implications for Reform*, 50 BUFF. L. REV. 103, 171 (2002) ("The experimental research suggests that caps [on damages] may have the counterintuitive effects of increasing both the size and variability of punitive damage awards in some cases.").

6. See Robbennolt, *supra* note 3, at 471 ("Compared to the extensive study of the decisionmaking of jurors and juries, there has been relatively little examination of trial

Before even reaching the question of the relative competence of judges and jurors, however, we should ask at least two questions about the research ostensibly indicating that a factfinder performs a task suboptimally: Are the research conclusions valid and generalizable to the settings of interest? Does the research utilize the proper test of optimal performance?⁷ When Brian Bornstein and Sean McCabe ask whether we should be concerned with the necessary use of hypothetical outcomes in trial simulations,⁸ they raise an important but often neglected issue that implicates both the validity and generalizability of experimental factfinder research. Furthermore, Professor Robbennolt notes the difficulty that can arise in choosing the proper test of optimality.⁹

The consequentiality issue addressed by Bornstein and McCabe raises construct validity concerns because mock jurors and judges deciding hypothetical cases with imaginary outcomes perform a task indisputably and inevitably different from the task of jurors and judges deciding actual cases with real consequences.¹⁰ Nevertheless,

judges' decisionmaking, and even fewer studies have directly compared the decisionmaking of juries and judges."). Robbennolt and other participants in the conference are working to correct this imbalance. See, e.g., Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777 (2001); Jennifer K. Robbennolt, *Punitive Damage Decision Making: The Decisions of Citizens and Trial Court Judges*, 26 LAW & HUM. BEHAV. 315 (2002).

7. Cook and Campbell divide questions about research validity and generalizability into four useful subcategories: (1) statistical conclusion validity, which concerns whether conclusions about statistically detected covariation among variables represent real or spurious relationships; (2) internal validity, which concerns whether conclusions about the causal relationship among variables represent valid conclusions; (3) construct validity, which concerns whether the constructs of interest have been properly operationalized for research purposes and whether generalizations about these constructs from the research setting to other settings can be validly made; and (4) external validity, which refers to the degree to which conclusions about the relationships among variables as found in the research setting can be generalized to other populations of persons, situations, and times. THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION: DESIGN & ANALYSIS ISSUES FOR FIELD SETTINGS* 37-39 (1979). For an interesting discussion of Campbell's development of the internal and external validity constructs, see Linda Albright & Thomas E. Malloy, *Experimental Validity: Brunswik, Campbell, Cronbach, and Enduring Issues*, 4 REV. GEN. PSYCHOL. 337, 340-43 (2000).

8. See Brian H. Bornstein & Sean G. McCabe, *Jurors of the Absurd? The Role of Consequentiality in Jury Simulation Research*, 32 FLA. ST. U. L. REV. 443 (2005).

9. See Robbennolt, *supra* note 3, at 502 ("One problem with comparing the decisionmaking of juries to that of judges is that to the extent that there are differences in the decisionmaking of judges and juries, there is not always a clear benchmark for determining which group's decision is normatively better.").

10. Another way to think of this construct validity concern is as an ever-present confound in the experiment: Is the mock jury's behavior attributable to the independent variable(s) or the lack of consequences?

The lack of consequences ranks with the lack of deliberation in trial simulations as the most pressing construct validity concerns in factfinder research. See David R. Shaffer & Shannon R. Wheatman, *Does Personality Influence Reactions to Judicial Instructions? Some Preliminary Findings and Possible Implications*, 6 PSYCHOL. PUB. POLY & L. 655, 657 (2000) ("In our opinion, perhaps the greatest limitation of mock-trial simulations is that the vast majority of them attempt to draw inferences from decisions rendered by non-

if the limited research comparing hypothetical and consequential decisionmaking uniformly found little or no difference between the consequential and inconsequential decisions, then we could dismiss this difference as insubstantial. Unfortunately, the research findings are mixed.¹¹ Furthermore, although the simulations used in jury research may prime the same cold cognitive processes primed in actual trials, the greater concern is that trial simulations fail to capture adequately the hotter cognitive processes involved in actual trials. Examples of the hotter cognitive processes include the emotions triggered by violent crimes and the stress of potentially erroneous decisionmaking in high-stakes cases, not to mention the personal and economic pressures on jurors that may arise from extended or highly publicized trials. Indeed, experimental designs typically strive to eliminate subject concerns about being held accountable for their decisions,¹² yet these concerns may play a powerful role in the actions of judges and jurors in actual trials.

It is ethically impossible to do laboratory research that would elicit the kinds of emotions and stress involved in many real cases, and it is next to impossible to receive approval from a human sub-

deliberating mock *jurors* rather than deliberating mock *juries*.”). The deliberation issue, however, has been investigated much more than the consequentiality issue. See generally Dennis J. Devine et al., *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 PSYCHOL. PUB. POLY & L. 622 (2001).

11. Bornstein and McCabe summarize the small and conflicting body of research on the consequentiality issue that has been conducted within the field of factfinder research. See Bornstein & McCabe, *supra* note 8, at 452-57. As Bornstein and McCabe note, there is some good evidence from research on framing effects that decisions with real and hypothetical consequences do not differ significantly, see *id.* at 462-65, but overall the evidence on incentives and consequences is mixed, see, e.g., Dan N. Stone & David A. Ziebart, *A Model of Financial Incentive Effects in Decision Making*, 61 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 250, 250 (1995) (“[E]vidence suggests that extrinsic incentives sometimes increase, sometimes decrease, and sometimes have no effect on decision quality.”). Thus, it is impossible to draw univocal conclusions about the effect of real consequences on decision processes. Cf. 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, 34 (1999) (“The data show that incentives sometimes improve performance, but often don’t. This unsurprising conclusion implies that we should immediately push beyond debating the caricatured positions that incentives always help or never help. Adopting either position, or pretending that others do, is empirically misguided and scientifically counterproductive.”).

12. See Karen Seigel-Jacobs & J. Frank Yates, *Effects of Procedural and Outcome Accountability on Judgment Quality*, 65 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 1 (1996).

Interestingly, although a great many real-world judgment and decision-making situations clearly include some level of accountability, most laboratory studies of judgment and decision-making are deliberately designed to minimize this factor. For example, subjects are generally told that any responses they make during the course of the experiment will be kept strictly confidential and anonymous, and there are seldom any real consequences for making “good” versus “bad” judgments or decisions.

Id. at 1-2.

jects committee to do experiments that would deceptively lead participants to believe they were making consequential decisions. It is also extremely rare to perform quasi-experimental factfinder research in court settings.¹³ For these reasons, the only viable solution to the construct validity problem raised by the consequentiality issue would seem to be a multimethod approach along the lines proposed by Bornstein and McCabe: supplement trial simulations with observational and survey research involving jurors and judges from real trials, utilize shadow juries in real cases, and perform experimental and quasi-experimental research on actual decisionmakers in administrative settings that may be more permissive of such research.¹⁴ Until we have convergence on the importance of the consequentiality dimension of factfinder decisionmaking across a range of settings and methods, the construct validity of decisions made in trial simulations will rightly remain highly suspect.¹⁵

The consequentiality issue raises external validity concerns because individuals may react differently in making hypothetical versus consequential decisions and because consequences of different magnitudes and types may have quite different psychological and behavioral effects both within and across individuals.¹⁶ While all hypothetical decisions have no consequences for the hypothetical parties and at most very limited social and self-evaluative consequences for the decisionmakers in these hypothetical cases, consequences for the parties in real cases can range from nominal damages to death penalties and pose much greater potential social and personal costs to the decisionmaker. A demonstration of external validity in trial simulations requires a demonstration that both the subjects and situations sampled in the experiments are representative of the fact-

13. The Arizona Jury Project involved the first experimental manipulation using real juries, with juries being randomly assigned to conditions in which discussion of evidence was or was not allowed before the close of evidence. The project also involved the taping of actual jury deliberations, which has happened in only three other instances. See Shari Seidman Diamond & Neil Vidmar, *Jury Room Ruminations on Forbidden Topics*, 87 VA. L. REV. 1857, 1867-71 (2001) (describing the Arizona Jury Project and discussing prior research involving jury deliberations).

14. See Bornstein & McCabe, *supra* note 8, at 465-66.

15. Note that the consequentiality/construct validity issue implicates many aspects of factfinder behavior, though its implications are most direct for ultimate decisions on liability and damages. For instance, jury attention to and comprehension of instructions could vary with the consequences of a trial.

16. On the latter point, Camerer and Hogarth note:

The data show that higher levels of incentives have the largest effects in judgment and decision tasks. Incentives improve performance in easy tasks that are effort responsive, like judgment, prediction, problem-solving, recalling items from memory, or clerical tasks. Incentives sometimes hurt when problems are too difficult or when simple intuition or habit provides an optimal answer and thinking harder makes things worse.

Camerer & Hogarth, *supra* note 11, at 34.

finders and trial situations to which the results are deemed inductively instructive.¹⁷

The external validity problem posed by the consequentiality issue is more tractable than the construct validity problem because we can examine a diverse range of subject and situational subpopulations in experimental settings. Moreover, the hypothetical stakes of simulated decisions may easily be varied to examine how judgments about evidence and decisions in cases shift with hypothetical consequences to determine the generality of behavior in simulations.¹⁸

Just as foundational as the questions about the validity and generalizability of trial simulations¹⁹ is the question of what are the proper normative criteria for assessing judge or jury competence, an issue which Robbennolt alludes to in her article.²⁰ Answering this question of proper normative standards presents a host of additional difficulties, some quite tractable and others resistant to easy solutions. First, the researcher must make a judgment about the goals that should be sought by a factfinder in performing a particular task, such as achieving truth or rationality in decisions about criminal guilt or civil liability. For instance, a jury's decision may be evaluated in relation to the empirical facts (did the convicted defendant actually commit the murder?) or in relation to an ideal treatment of evidence presented at trial in light of the rules of evidence and the rules of statistical inference (did the jury draw rational inferences from the evidence presented at trial in light of the jury's instructions?).²¹ The choice of proper goals is crucial because different goals,

17. Egon Brunswik, *Representative Design and Probabilistic Theory in a Functional Psychology*, in *THE ESSENTIAL BRUNSWIK: BEGINNINGS, EXPLICATIONS, APPLICATIONS* 135, 140 (Kenneth R. Hammond & Thomas R. Stewart eds., 2001) ("As we cannot possibly hope to encompass the entire population of individuals in research, but must sample representatively, we must sample instances in the study of functional achievement.").

18. Interestingly, Kühberger and his colleagues, in their study of framing effects in real and hypothetical gambles, found similar effects of the magnitude of consequences for both real and hypothetical decisions, which suggests that trial simulations may usefully examine hypothetical cases with a range of consequences. See Anton Kühberger et al., *Framing Decisions: Hypothetical and Real*, 89 *ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES* 1162, 1173 (2002) ("In conclusion, we found that the size of the incentive influenced participants' choices. Paradoxically, for 'incentive theory,' the high incentive has the same effect in the real and hypothetical gambles.").

19. Bornstein & McCabe, *supra* note 8.

20. See Robbennolt, *supra* note 3, at 502.

21. Friedland emphasizes accuracy as the goal of juror decisionmaking, see Steven I. Friedland, *The Competency and Responsibility of Jurors in Deciding Cases*, 85 *NW. U. L. REV.* 190, 195 (1990) ("[A]ccurate juror decisionmaking is a predicate to a public perception of fair decisionmaking."), while Diamond emphasizes adherence to the law given the difficulty of determining accuracy:

To assess how the jury operates as a decision-maker, we cannot compare the jury's verdict with some gold standard of truth because no such dependable standard exists. . . .

or different norms for implementing those goals, may point to different conclusions about competence.²² A jury presented with inadequate evidence at trial may irrationally convict the truly guilty defendant, while another jury presented with this inadequate evidence may rationally acquit the truly guilty defendant. The first jury did a good job under the correspondence standard but a poor job under the coherence standard, and conversely, the second jury performed well under the coherence standard but poorly under the correspondence standard. Hence, the choice of normative criteria controls the assessment of the factfinder's competence on particular tasks, and the choice of goal determines the choice of normative criteria.²³

Second, because there may be disputes about the goals that should be served by certain tasks, the researcher must employ multiple, possibly conflicting normative criteria or justify her choice of

We can, in contrast, say something about the quality of juror decision-making by assessing whether the jurors have considered relevant facts and legal instructions in arriving at their verdicts.

Shari Seidman Diamond, *Truth, Justice, and the Jury*, 26 HARV. J.L. & PUB. POL'Y 143, 150-51 (2003).

The first approach, which emphasizes accuracy, follows from a correspondence metatheory of good judgment and decisionmaking: "The goal of a correspondence metatheory is to describe and explain the process by which a person's judgments achieve *empirical accuracy*." Kenneth R. Hammond, *Coherence and Correspondence Theories in Judgment and Decision Making*, in JUDGMENT AND DECISION MAKING: AN INTERDISCIPLINARY READER 53, 53 (Terry Connolly et al. eds., 2d ed. 2000). The second approach, which emphasizes adherence to rules, follows from a coherence metatheory of good judgment and decisionmaking: "The goal of a coherence metatheory of judgment . . . is to describe and explain the process by which a person's judgments achieve logical, or mathematical, or statistical *rationality*." *Id.*

22. Some goals, such as accuracy in decisionmaking, require no more specific implementing norms to assess competence: the decision is simply reviewed for its correspondence to the empirical facts. Other goals, such as rationality in decisionmaking, require more specific norms to implement the goal and conduct the competence evaluation. For instance, within economics, rational choice typically means the maximization of expected utility or subjective expected utility, and within this view the maximization of utility requires fidelity to axioms put forward by von Neumann and Morgenstern or Savage or variants thereof. See generally Peter C. Fishburn, *Decision Theory: The Next 100 Years?*, 101 ECON. J. 27 (1991); Paul J.H. Schoemaker, *The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations*, 20 J. ECON. LITERATURE 529 (1982). Within the legal setting, rational factfinding requires adherence both to applicable legal rules and rules of probability for evaluating evidence under conditions of uncertainty or ambiguity. As discussed below, choosing these implementing norms may present additional complications.

23. Discussions of factfinder competence typically employ external measures of competence, rather than internal measures that ask whether the factfinder achieved its own self-chosen goals. Because we are interested in how factfinders fulfill their roles within the larger legal-institutional setting, it is fair to impose values and goals on the factfinder that the factfinder might not endorse or that might even be contrary to the factfinder's own goals or values. A finding of external incompetence thus does not mean that the factfinder was incompetent from an internal perspective; it only means that the factfinder failed to meet an externally imposed normative standard. Indeed, a factfinder may be deemed externally incompetent precisely because the factfinder is quite competent at achieving a goal that is internally desirable but undesirable from an institutional, external standpoint.

some normative criteria to the exclusion of others.²⁴ Otherwise, the research may be dismissed as irrelevant or disputed on grounds of incompleteness.²⁵

Third, even if there is no dispute over the proper goal to be served by a particular task, there may be dispute over the proper normative criteria for judging whether that goal has been achieved. For instance, when juries make judgments about the reasonableness of a party's actions, one of the goals, and perhaps the primary goal, is to inject objectivity into the proceedings and transcend the subjective perceptions of the parties.²⁶ Yet it is not at all clear how best to assess a jury's competence at being reasonable and objective.²⁷ Another

24. Or it may be that different legal systems embrace different goals, and it may be impracticable to consider all of the different goals within a single study if each goal requires a variation in design to test for optimal performance.

25. We see some debate of this nature within the field of punitive damages research. For instance, although Professor Sharkey disputes some of the details of the empirical work by Sunstein and others that questions the coherence of punitive damages awarded by juries, Sharkey's primary difference with Sunstein and his colleagues seems to be that she prescribes different purposes for punitive damages than Sunstein and his colleagues. See Catherine M. Sharkey, *Punitive Damages: Should Juries Decide?*, 82 TEX. L. REV. 381, 385 (2003) (reviewing CASS R. SUNSTEIN ET AL., PUNITIVE DAMAGES: HOW JURIES DECIDE (2002)) ("A fuller examination of Sunstein et al.'s empirical work reveals the indeterminacy—and possible inapplicability—of that research with respect to broader nonretributive theories of punitive damages."); *id.* at 412 ("[E]ven if Sunstein et al. may have established the inability of jurors to translate moral outrage into dollar awards, they have not sufficiently explored the jurors' assessment of deterrence-based or socially compensatory damages, wholly separate, but by no means exclusive, categories of punitive damages."); see also Neal R. Feigenson, *Can Tort Juries Punish Competently?*, 78 CHI.-KENT L. REV. 239, 284-88 (2003) (reviewing CASS R. SUNSTEIN ET AL., PUNITIVE DAMAGES: HOW JURIES DECIDE (2002)) (arguing that Sunstein and his colleagues fail to justify their emphasis on optimal deterrence as the goal of punitive damages to the exclusion of retribution as a goal).

26. See, e.g., Brian Leiter, *Introduction to OBJECTIVITY IN LAW AND MORALS* 1, 3 (Brian Leiter ed., 2001) ("In some areas of law, we expect the law to employ 'objective' standards of conduct (like 'reasonable person' standards) that do not permit actors to excuse their conduct based on their subjective perceptions at the time."). For a discussion of several possible interpretations and operationalizations of the reasonableness construct, see Neil MacCormick, *Reasonableness and Objectivity*, 74 NOTRE DAME L. REV. 1575, 1579-83 (1999).

27. Even within negligence suits, where one might expect reasonableness to mean engaging in a particular course of action only when the expected benefits outweigh the expected costs, juries typically are not instructed to perform such a balancing analysis when applying the reasonable-person standard. See Steven Hetcher, *The Jury's Out: Social Norms' Misunderstood Role in Negligence Law*, 91 GEO. L.J. 633, 639 (2003) ("In negligence suits, the jury instruction regarding liability is typically stated in terms of the reasonable person standard. Usually there is little further instruction telling the jury how to interpret this standard or determine the meaning of the crucial concept of reasonableness." (footnote omitted)); see also Mark P. Gergen, *The Jury's Role in Deciding Normative Issues in the American Common Law*, 68 FORDHAM L. REV. 407, 408 n.1 (1999) ("It is a gross misunderstanding of the law of negligence to claim that the standard of the reasonably prudent person can be reduced to the Hand formula without losing something vital."); Patrick J. Kelley & Laurel A. Wendt, *What Judges Tell Juries About Negligence: A Review of Pattern Jury Instructions*, 77 CHI.-KENT L. REV. 587, 618 (2002) ("[T]he cost-benefit test of negligence

pertinent example of indeterminacy in normative criteria involves the proper weight to be given to incremental bits of evidence to form rational judgments about guilt or liability and, in particular, whether Bayes' Theorem offers the best way to adjust beliefs in light of new evidence.²⁸ In such cases, little can be done but to stick stubbornly to one's preferred normative view or acknowledge the debate over proper norms and apply multiple criteria.

A fourth obstacle arises when researchers must translate vague goals into specific norms for testing purposes.²⁹ If the goal is to achieve coherence or efficiency in outcomes, for example, then the researcher must come up with a way to measure coherence or efficiency. Experiments provide a particularly attractive setting for addressing this problem because they allow the researcher to vary the dimensions of hypothetical cases to examine whether factfinder behavior conforms to different formulations of the vague goals. In addition, experiments require that the researcher commit beforehand to particular theoretical formulations of coherence, efficiency, or other vague goals, thus avoiding the biasing effect that preexisting views may have on the search for and interpretation of evidence when trying to test for coherence or efficiency with actual historical cases.³⁰

does not seem to be the probable meaning of even those five pattern negligence instructions couched in terms of unreasonable foreseeable risk.”).

28. See Alvin I. Goldman, *Quasi-Objective Bayesianism and Legal Evidence*, 42 JURIMETRICS J. 237, 238 (2002). For example, Goldman notes:

Bayesianism is a popular approach to legal evidence. Entire books have been devoted to the analytical power of Bayesian inference and its application to legal evidence. At the same time, many doubts and challenges have been raised to the adequacy of Bayesianism in general and its specific application to the law.

Id. (footnotes omitted). For a recent example of opposition to Bayesianism, see Stuart E. Thiel, *Probability Models of Juridical Proof: It's Time to Kick Bayes Out on His Posterior 2* (June 8, 2004) (unpublished manuscript) (“My thesis is that the Bayesian model poorly fits the problem of juridical proof, and that the statistical decision paradigm, ‘frequentist’ or ‘classical’ statistics, is better-suited for the task and does a better job.”), available at <http://ssrn.com/abstract=556028> (last visited Dec. 13, 2004).

29. In some cases, it may simply be impossible to formulate workable tests of goals, which may result in the selection of a second-best goal. Indeed, we might prefer that truth be the ultimate touchstone in factfinding, but we may have to settle for a coherence measure of competence over a correspondence measure because we lack a reliable measure of truth in many cases. Of course, while many of the rules of evidence can be seen as rules designed to lead to accurate decisionmaking, other rules endorse values other than truth and prevent the introduction of evidence that might improve accuracy. We may thus favor a coherence standard that asks whether juries follow the rules of evidence because we believe that these rules are the most feasible and manageable means to truth approximation in light of epistemic constraints and because of the other values contrary to truth detection that the trial process must serve.

30. Legal scholarship often employs interpretive analyses of caselaw to support various propositions, yet such interpretive analyses raise serious selection bias problems because of the unrepresentative samples of caselaw typically used. For instance, Christopher Bruce reviewed articles published in *The Journal of Legal Studies* between 1972 and 1988 and found little evidence of the use of random sampling of cases in law and economic stud-

The final component in the examination of factfinder incompetence is the search for causal mechanisms: once we apply a norm and find suboptimal behavior through valid and reliable research methods, we should ask why the factfinder fails to achieve the goal in question, in hopes of improving performance or learning whether the goal is unrealistic. As Robbennolt notes, judges and jurors occupy drastically different positions in terms of decision process, knowledge, and other resources.³¹ By better understanding whether judges and juries differ in their behavior, and if so, why, we may begin to exploit the comparative advantages of the judge and jury or refashion procedures to eradicate any unwanted differences and better achieve the desired goals through either factfinder. Indeed, rather than assume that judges or juries will do a better job than the other at some particular task when suboptimal behavior is observed in one or the other, we should engage in task analysis to understand why the task was so difficult in the first place. It may be that no judge or jury should be expected to perform a task optimally given the demands or constraints of the situation, in which case it is the task that needs change rather than the assignment of the task to judges versus juries.³²

ies published in that journal during that period. See Christopher J. Bruce, *A Positive Analysis of Methodology in the Law and Economics Literature*, 12 *HAMLIN L. REV.* 197, 219 (1989) ("It was found that random samples were rarely used in the discussions of both tort and contract law. Random sampling was found to have become more prevalent over time, but not to have dominated non-random sampling."). Recently, Penelope Pether made the more radical argument that published American caselaw in general provides an incomplete and biased picture of the law as it really exists and is applied. Penelope Pether, *Inequitable Injunctions: The Scandal of Private Judging in the U.S. Courts*, 56 *STAN. L. REV.* 1435, 1441 (2004) ("[T]here is credible evidence of the tendency for the practices of private judging to corrupt the operation of the courts and the administration of justice . . ."); see also David S. Law, *Strategic Judicial Lawmaking: Ideology, Publication, and Asylum Law in the Ninth Circuit*, 73 *U. CIN. L. REV.* (forthcoming 2005) (available in an earlier version as David S. Law, *Strategic Judicial Lawmaking: An Empirical Investigation of Ideology and Publication on the U.S. Court of Appeals for the Ninth Circuit*, at http://papers.ssrn.com/abstract_id=602861 (last visited Feb. 1, 2005)) (arguing that ideology may bias judicial decisions about which cases are published versus unpublished). One need not agree with Pether's argument, however, to recognize that published caselaw is a poor sample for examining the competence of judges and juries given the paucity of important details (especially in jury trials), the disparities among trial judges in propensity to publish opinions, the relative paucity of reported decisions from state trial courts as compared to federal courts, and the difficulty of establishing whether published cases fairly represent the universe of cases.

31. See Robbennolt, *supra* note 3, at 502-06.

32. David Schkade provides an excellent example of the value of taking this task design perspective once suboptimal behavior is observed, and the conclusion from his analysis of the task of setting punitive damages is worth repeating here:

If jury punitive damage awards surprise us and seem unpredictable, it is due more to the situation that jurors find themselves in than to the characteristics of the jurors themselves. Several features of the task of assessing an amount of punitive damages seem almost designed to produce erratic awards. Unfortunately, it is not clear that these flaws can be corrected while retaining the full

