

---

Research Note

---

## Illuminations and Shadows from Jury Simulations

Shari Seidman Diamond<sup>1,2</sup>

---

*In the 18 years since authors in Law and Human Behavior's special issue on simulation research called for more realistic jury simulations, jury simulations of all kinds have proliferated. While simulations in general represent a significant improvement over nonempirical armchair speculation about jury behavior, the more ecologically valid features of recent simulations increase both the quality and the persuasiveness of simulation results. Still missing, however, are theories and a data base that will signal when these more elaborate and expensive design features are crucial.*

---

### INTRODUCTION

How well do jury simulations reflect jury decision-making? What methodological features affect the ability of simulations to provide trustworthy information about juries? How are courts likely to respond to simulations that purport to represent how juries behave? What can researchers interested in policy do to ensure that courts are provided with valuable and persuasive evidence from simulations?

In 1979, *Law and Human Behavior* published a special issue on Simulation Research and the Law (Diamond, 1979). Authors writing in that issue raised questions about the theoretical, empirical, and policy value of available research on the jury. Most of the jury experiments in the literature at that time were based on rather crude and minimal trial simulations. Some of us (e.g., Vidmar, 1979; Weiten & Diamond, 1979) suggested strategies that we believed could improve the caliber of jury research (e.g., using videotaped trials instead of short, written vignettes as experimental stimuli; studying juror or community-member respondents rather than students from college subject pools; using more ecologically valid dependent measures).

<sup>1</sup>American Bar Foundation, Chicago, IL, and Department of Psychology, University of Illinois, Chicago, IL.

<sup>2</sup>Correspondence should be addressed to Shari Seidman Diamond, American Bar Foundation, 750 N. Lake Shore Drive, Chicago, IL 60611.

The jury research literature has continued to grow during the past 18 years, and a number of the studies published during that period have incorporated the recommended strategies. For example, Hastie, Penrod, and Pennington (1983) studied the effects of decision rule using jury venire members who viewed a videotape of a homicide case, and others (e.g., Cowan, Thompson, & Ellsworth, 1984; Ellsworth, 1989) have used the Hastie et al. tape to study a variety of issues. Brekke, Enko, Clavet, and Sellau (1991) showed jury-eligible citizens a videotaped reenactment of a sexual assault trial. Jonathan Casper and I have examined the reactions of jury venire members to a videotaped antitrust trial on damages (Diamond & Casper, 1992) and to a death penalty hearing (Diamond, Casper, Heiert, & Marshall, 1996). I am currently analyzing the experimental results from a third large-scale jury simulation using a videotaped products liability trial.<sup>3</sup> In conducting each of these simulations, the authors created multiple versions of the trial to vary aspects of the testimony or instructions so that the effect of those variations on juror behavior could be assessed. In our antitrust, death penalty, and products liability cases we used professional actors in a courtroom setting. We showed the antitrust and death penalty trials to jurors awaiting assignment to trials in Cook County courthouses, and the products liability trial to jury-eligible community members whose demographics match the local jury venire. All of this is expensive and time-consuming work, so it is worth asking what is gained by such efforts. Have these more elaborate attempts at verisimilitude been worth the trouble?

Some of this research has been presented to trial courts and in appellate briefs, giving judges an opportunity to comment on methodological characteristics that enhance or detract from its persuasiveness (e.g., *Free v. Peters*, 1993; *Lockhart v. McCree*, 1986). As a result, we also have further evidence on the reactions of legal decision-makers to this type of evidence. What have we learned?

### AN EVALUATION

Weiten and Diamond (1979) identified six methodological characteristics typical of existing jury simulation studies that might reduce the validity of conclusions drawn from that research: inadequate sampling, inadequate trial simulations, lack of jury deliberation, inappropriate dependent variables, lack of corroborative field data, and the nature of decisions based on role-playing. Researchers have dealt with some, but not all, of these characteristics in a number of recent studies. The results provide some insights on the effects of these features on what simulation research can tell us about jury behavior.

<sup>3</sup>My collaborator on the first two trials was Jonathan Casper, and my collaborators on the third are Stephen Landsman, Michael Saks, and Linda Dimitropoulos.

### **Inadequate Sampling**

Increasingly, researchers have turned to courthouse jury venires and community samples for respondents, using student populations for preliminary work. Although some research has shown parallel results for student and jury pool samples (e.g., Casper, Benedict, & Perry, 1989), other research indicates that even subsets of former jurors who agreed to participate in research conducted after jury service may differ in significant ways from jurors who become research participants in the course of jury duty. The volunteer ex-jurors that Severance, Greene, and Loftus (1984) recruited were older and more educated than the current jurors they studied, and the verdicts of the two groups and the impact of deliberations on their verdicts were also different. In developing a videotaped death penalty hearing, half of the college students Casper and I tested voted in favor of imposing the death penalty, but we had to revise the hearing transcript to obtain the same split on verdict preference from the members of the jury venire: jurors were much more inclined than the students to favor imposing the death penalty on the defendant in the original version of the case.

Sears (1986) identified a number of attributes of college students that may make it hazardous to use their behavior on cognitive tasks or in situations involving conformity pressures to study how a general adult population would respond. College students are particularly susceptible to normative pressures, especially from peers, and are more accomplished at cognitive tasks than many members of the general population. Thus, for example, simulations with college students may overestimate the ability of jurors to handle complex trial testimony or instructions and the power of an initial majority to produce a unanimous verdict.

If differences between student and general adult populations were confined simply to differences in rates of conviction or willingness to find liability, and did not interact with experimental manipulations, the choice of a subject population would not be troubling. The difficulty is that we have not yet developed and tested any theory that will tell us when differences will and will not emerge. In light of the limited and inconsistent comparative tests currently available, two forms of data collection seem appropriate. First, to build a data base that will allow us to examine systematically the differences between student and adult samples, jury researchers should routinely conduct Stage One work on college students that they can compare with Stage Two results obtained from respondents more representative of jury venires. Second, until we can predict reliably when a student sample is likely to provide an adequate model of juror behavior, jury venires should remain the preferred source for maximizing both face and external validity.

### **Inadequate Trial Simulations**

Trials are complex time-consuming events, and even the most elaborate simulations last no more than a few hours, truncating the contents and the time sequence of a normal trial. They are, however, far more lengthy and complex than the 300-word case summary that typified early jury simulations. Both

theoretical predictions (Kaplan & Miller, 1978) and some empirical evidence (Baumeister & Darley, 1982; Linz & Penrod, 1982) suggest that experimental effect sizes will be magnified when they appear in the context of abbreviated trial simulations. Potential distortions are particularly likely in some of the work on the effects on jury decision-making of defendant or victim characteristics such as race and gender (e.g., McGlynn, Megas, & Benson, 1976; Ugwuegbu, 1979). All decision-makers are most likely to rely on stereotypes when they are given little additional information on which to base decisions (Hamilton & Sherman, 1996). Thus, a skeletal case description maximizes the likelihood that stereotypes will be used to complete the picture. In contrast, a more extensive familiarity with a defendant or witness gives the decision-maker other grounds on which to base a decision. Because actual trials inevitably provide a myriad of additional legally relevant and irrelevant bases on which to make a decision, effects of race and gender may be substantially modified when jurors are provided with that additional contextual information.

Nonetheless, more abbreviated experimental stimulus materials *can* play an important role in addressing some questions about jury behavior. Suppose, for example, that an experimenter believes that jurors will not only be more likely to find a defendant liable, but will also give larger damage awards if testimony reveals that the defendant had prior knowledge about the potential danger of a broken stairwell where an accident occurred. If a strong presentation of information on prior knowledge produces no effect on damage preferences in a small vignette simulation, it is unlikely that the information about prior knowledge will be a powerful enough influence to affect jurors if it is embedded in a full-scale mock trial in which it may not receive particular emphasis.

A less-than-full-scale trial stimulus can also provide a suitable vehicle if the research is designed to study decision-making issues that are not dependent on the nature of the evidence being presented at trial. For example, while a study of the credibility of various types of expert testimony may require a fairly elaborate simulation to capture the relevant experience of a juror facing such testimony, a less extensive trial stimulus might be sufficient for a study designed to test the effect of the unanimity requirement on deliberations.

Some researchers have not observed differences between reactions to abbreviated and more extensive trial stimulus materials (Kramer & Kerr, 1989), and there is no doubt that creating such materials is an expensive and time-consuming process. Nonetheless, in view of the evidence that such differences sometimes do occur, the trend toward more realistic trial simulations appears to be justified.

### **Lack of Deliberations**

The prevailing view, which originated with Kalven and Zeisel (1966), is that deliberations play a minor role in determining jury verdicts because the predeliberation majority generally prevails in the end. Kalven and Zeisel (1966) came to this view based on interviews with jurors who reported the first ballot

distribution of their jury and the jury's verdict. They suggest that "with very few exceptions the first ballot decides the outcome of the verdict" and that therefore "the real decision is often made before the deliberation begins" (p. 488). Their conclusion, based on the first ballot evidence, is warranted only if the first ballot reflects the distribution of predeliberation juror verdict preferences unaffected by persuasion that has already occurred in deliberations. Some recent work suggests that significant discussion often precedes any vote by the jury (Sandys & Dillehay, 1995). Thus, the strong relationship between first ballots and verdict revealed by Kalven and Zeisel may fail to disclose considerable opinion shifts that take place as a result of deliberations, but before a first vote is taken.

While a number of researchers have shown a strong relationship between predeliberation distribution of verdict preferences on a jury and the jury's verdict (e.g., Stasser, Kerr, & Bray, 1982), a small but growing number of studies that have included deliberations indicate that deliberations sometimes do influence outcomes. MacCoun and Kerr's (1988) research revealed a leniency shift associated with criminal cases that use a reasonable doubt standard. Kerwin and Shaffer (1994) found that student jurors who participated in deliberations were more likely to follow judicial instructions to ignore inadmissible testimony than jurors who responded individually without deliberating. Other researchers have found evidence that deliberation can reduce other biases that operate at the individual juror level (Kaplan & Miller, 1978). In general, it appears that deliberations can produce jury verdicts more consistent with legal norms, but other evidence suggests that this effect may be limited to situations in which the legal instructions present those norms clearly.

One potential value of all groups, including juries, is that they provide an opportunity to pool information and correct misunderstandings. Thus, if jurors effectively share information and correct one another's errors, comprehension should improve following deliberations. Jurors in at least one study showed greater comprehension of trial testimony, but not judicial instructions on the law, following deliberations (Ellsworth, 1989). In another study, jurors showed a significant improvement in comprehension of legal instructions, but only when a substantial majority began deliberations with a correct understanding of the relevant instruction (Diamond & Levi, 1996). This pattern suggests that deliberations will reduce error rates only if a significant proportion of the jurors begin deliberations with correct information; otherwise, deliberation may simply reinforce the inaccuracies of the majority.

Taken as a whole, these results point to the value of testing the impact of deliberations on jury verdicts. Although juries are made up of jurors, and the patterns of response obtained from studies of individual jurors will generally predict jury outcomes, our recent findings suggest that deliberations do not simply develop the photographic negative present in the predeliberation vote distribution; careful studies of the deliberation process will thus add to our understanding of the jury.

### Inappropriate Dependent Measures

The standard use of degree of responsibility and years in prison as the primary dependent variables in jury research has all but disappeared. Breaking the hold of these psychometrically attractive continuous variables in favor of the more realistic, but less sensitive, dichotomous or at least categorical verdicts delivered by juries in most states has been a major corrective to the kind of legal naivete that Vidmar (1979) complained about 18 years ago. An interesting adjunct to the dichotomous guilty-not guilty verdict has been an index that combines verdict choice with a scale of confidence. Thus, Kassin and Wrightsman (1979) created an index that ranged from -8 to +8 by multiplying juror confidence ratings (on an 8-point scale) by 1 for jurors who favored a guilty verdict and by -1 for jurors who favored a not guilty verdict. This type of index generally has a U-shaped distribution, so that it will not always be more sensitive than the dichotomous measure. We do, however, have some evidence of the measure's ecological validity. In our death penalty study, Jonathan Casper and I measured juror verdict preferences and confidence ratings before and after deliberations. Whether a juror initially favored or opposed a death sentence for the defendant, jurors who changed their verdict choice following deliberation had expressed significantly less confidence in their predeliberation verdict choice than jurors who did not change their verdict preference. Thus, the confidence a juror expressed in an initial verdict preference was an indicator of how strongly the juror held that view, and the kind of combination index used by Kassin and Wrightsman (1979) reflects this position.

### Role Playing and Corroborative Field Data

The two remaining characteristics we identified in 1979 as potential weaknesses in simulation research endure, despite our ever more elaborate attempts at verisimilitude. Field studies have accumulated, albeit unsystematically, and while they give little indication of inconsistency between laboratory and field, few topics have been the subject of parallel research in both settings [for an exception, see Hans and Lofquist's (1992) study of cases involving corporate defendants]. More importantly, no further studies with jurors on the effect of role playing have been done since the research we reported on 18 years ago [but see an interesting study by Kaplan and Krupa (1986), who told student jurors that they were acting as real judges in an actual case of alleged student cheating or as simulated judges].

In summary, then, we can take some comfort that our efforts to invest more resources in the jury simulation paradigm have been justified. Nonetheless, we must acknowledge that because the costs of this approach are substantial, they impose tradeoffs. Moreover, even these more elaborate simulations cannot avoid some of the inevitable uncertainties that research can reduce but not avoid.

First, any simulation that depends on a single case is potentially idiosyncratic in ways that may modify the influence of independent variables being studied. The strategy of studying multiple cases is easy to act on when a one-page vignette is used. When an expensive videotaped trial must be created, researchers are less likely to test for consistent

results across cases. Inevitable questions about generalizability arise, however, if we use the same case over and over again to test all of our hypotheses about the jury.

Second, we need to develop and test a theory that can accurately predict the circumstances under which real consequences of a decision do and do not affect the pattern of decision-making. Respondents in a simulation often exhibit patterns of behavior observed outside the laboratory setting, including anger, frustration, and confusion, but behavior in the course of a simulation provides no direct evidence that it provides an accurate reflection of jury activity.

What then distinguishes the real decision for the respondent? Presumably, a juror in a real case is more motivated to attend to the trial evidence, to determine the verdict he or she really prefers, and to persuade fellow jurors that the jury verdict is consistent with that preference. Yet, assuming that real jurors do have a higher level of motivation, their heightened motivation may affect only some of the behaviors that psychologists investigate. For example, researchers have used simulations to study the influence of the hindsight bias on juror decisions (Casper et al., 1989). Because of the hindsight bias, individuals asked to judge the foreseeability of an outcome tend to see it as more likely if they are made aware of how events actually turned out (Fischhoff, 1975). Motivation may have little to do with the hindsight bias. Connolly and Bukszar (1990) found that respondents who did and did not have a stake in predicting the case outcome displayed the hindsight bias equally. It appears that the mechanism of the bias is cognitive rather than motivational. In such cases, concern about applicability to decision-making with real consequences may be overstated. Similar work testing the effect of incentives on behavior needs to be done in other areas. Without serious theoretical development and research on the consequences of consequences, we must be circumspect about the predictive value of our simulations.

Yet despite these remaining questions about even the most elaborate simulations, it is important to evaluate them in the context of other ways of testing assumptions about jury behavior. Imagine the armchair alternative to the extensive simulation work done in preparing for a space walk. Traditional nonscientific and nonempirical ways of knowing are likely to be far less trustworthy sources of information than are carefully done simulations.

### COURT REACTIONS

Even if researchers are prepared to believe that simulation findings provide valuable insights about jury behavior, courts need not agree with that evaluation. Of course researchers may choose to study jury behavior without expressing any interest that their work should influence policy. Scholars who believe that their insights have policy implications, however, may be troubled by recent responses from some members of the judiciary. The extensive body of jury simulation evidence presented to the U.S. Supreme Court in *Lockhart v. McCree* (1986) drew a negative reaction from a five-member majority.<sup>4</sup> Justice Rehnquist found that only 6 of the 15 studies that were presented dealt with the central issue in the case, namely, the potential effects

<sup>4</sup>Empiricist Justice Blackmun concurred with the result, but not the opinion of Justice Rehnquist.

on the determination of guilt or innocence of excluding from the jury individuals whose opinions on the death penalty would disqualify them from determining whether a defendant would be sentenced to die. He eliminated 3 of these 6 studies on the grounds that they had been considered “tentative and fragmentary” when presented in *Witherspoon v. Illinois* (1968) and were unchanged since that time.<sup>5</sup> Concern here is with the three remaining post-*Witherspoon* simulation studies that Justice Rehnquist ultimately set aside on methodological grounds.

Rehnquist criticized all three studies because they did not involve actual jurors deciding the fate of an actual capital defendant, and he expressed doubt that such studies could predict the behavior of actual jurors. It is unclear whether his doubts were stimulated primarily by the respondent characteristics, by the lack of decision consequences, or by both. It is also unclear whether he would have deemed these grounds sufficient to justify disregarding them. But he did not rely solely on these two attributes in rejecting the evidence. He also found two of the three studies objectionable because they failed to simulate deliberations. He found the same two studies fundamentally flawed because they failed to exclude individuals who would be ineligible to serve in the guilt phase due to their deep-seated opposition to the death penalty which would prevent them from deciding guilt or innocence fairly and impartially. At the end of the day, one study remained (i.e., Cowan, Thompson, & Ellsworth, 1984), criticized but apparently not discarded for its simulation nature and lack of actual jurors. Rehnquist concluded that a rule “as far reaching as the one McCree proposes should not be based on the results of the lone study that avoids [the] fundamental flaw [of including “nullifiers”]” (p. 172-173).<sup>6</sup>

In a more recent consideration of simulation research, the Seventh Circuit (*Free v. Peters*, 1993) dealt with a study in which actual members of the jury venire heard and read the description of a capital sentencing hearing and answered written questions aimed at gauging their comprehension of death penalty instructions. The federal district court had found, based on the empirical evidence, that the jury instructions provided constitutionally inadequate guidance to the jurors in reaching capital decisions. On appeal before the Seventh Circuit, Judge Posner (writing for a 2-1 majority) overturned the district court’s decision on procedural grounds and went on to criticize the research. He found the particular study questionable on two grounds. First, he characterized the procedures used as a written examination and opined that jurors might be quite capable of performing in a trial situation even if they performed poorly as test-takers responding to a written case summary that failed to capture the characteristics of a full sentencing hearing. Second, he complained that the research had failed to demonstrate that a revised instruction would produce a higher level of comprehension.<sup>7</sup>

<sup>5</sup>In fact, two of the three studies were later published. Only a preliminary report of the Zeisel (1968) study had been available to the Court in *Witherspoon*. The Goldberg study was published in 1970.

<sup>6</sup>Ellsworth (1988) has written eloquently about Justice Rehnquist’s “and-then-there-were-none” approach to the research evidence in *Lockhart* and his apparent insensitivity to the value of scientific triangulation of evidence from multiple albeit imperfect sources.

<sup>7</sup>The Constitution requires adequate guidance from judicial instructions in capital cases, so it is not clear why a *defendant* must supply an alternative instruction. Nonetheless, courts are more likely to acknowledge a problem when they are also provided with a solution.



Neither of these criticisms is fundamentally unreceptive to jury simulation research. In fact, Judge Posner suggested two simulation studies that would potentially address his concerns. The first study he suggested would compare performance on the test after seeing a mock sentencing hearing with performance by those who had seen only the written summary of evidence and instructions. Presumably, if jurors did not show better performance after watching a more extensive hearing simulation, their low performance in response to the more abbreviated materials could not properly be attributed to that summary form. The second study he suggested, in which the researcher compares juror performance in response to original and rewritten instructions, has in fact been done. Linguist Judith Levi and I (Diamond and Levi, 1966) tested the comprehension of juror-eligible citizens to the originally tested instructions versus a simplified revised set of instructions. As others had shown earlier with non-capital jury instructions (e.g., Elwork, Sales, & Alfini, 1982), jurors who received the revised instructions performed significantly better, so the poor performance in the original study cannot be explained away by the claim that jurors simply are not good test-takers.

Are courts then irrevocably unfriendly to simulation research? The larger picture suggests they are not — or at least that there is enormous variation in response. Some courts have welcomed simulation studies, endorsing changes not only in jury instructions (*Mitchell v. Gonzales*, 1991), but also in rules of evidence (*People v. Allen*, 1988), and in procedures governing the conduct of jury trials (Arizona Supreme Court Committee on More Effective Use of Juries, 1994). Courts in general, particularly in a post-*Daubert* world (*Daubert v. Merrell Dow Pharmaceuticals*, 1993), are paying more attention to scientific evidence of all kinds. Walker and Monahan (1996) suggest that “Simulation research . . . is likely to receive greater judicial acceptance in the wake of *Daubert* and [the Federal Judicial Center’s 1994] *Reference Manual [on Scientific Evidence]*” (pp. 854–855).

Are courts skeptical about the weight they should give simulation research? Absolutely. Moreover, a court confronted with even the most elaborate simulation with optimal verisimilitude in the experimental stimulus, selection of respondents, and the inclusion of deliberation can always (reasonably) point to the fact that the trial is not real and that the jurors know that their decision will not affect the fate of the parties. Of course, we might (reasonably) suggest that the appropriate question is how probative the simulation evidence is when compared with other available evidence and that the simulation, despite weaknesses, offers more reliable and valid evidence than the judge’s empirically untested assumptions about jury behavior.

We might also offer a different sort of response. If courts are willing to entertain questions about jury functioning, how, we may ask, should these questions be addressed? Field experiments on real trials can be mounted when the questions concern procedures like permitting note-taking or question-asking (Heuer & Penrod, 1994), but more fundamental questions may require interventions that courts and parties will be unwilling to endorse (e.g., if we wished to test whether jurors are more likely to accept uncritically the testimony of a court-appointed versus party-appointed expert, could we manipulate this feature in a set of real trials?). Studies of jury verdict data, while attractive because they are often based on large

samples of real jury verdicts, face serious limitations due to the many plausible rival alternative hypotheses that cannot be eliminated (Vidmar, 1994). Even the most perceptive observer is unable to obtain trustworthy evidence about jury behavior if that analyst is forced to rely solely on postverdict interviews with jurors whose impressions of their initial reactions to the evidence and of the content of deliberations are likely to be transformed in the course of deliberations and in light of the verdicts they ultimately reached.

In the end, however, it is incumbent on us, if we want to assure ourselves and to persuade courts how informative our simulation research can be, to continue to test the effects of variations in methodology on our findings. Indifferent courts are likely to be more interested if we can tell them when and how our microscopes provide unique and accurate images not visible by other means.

### ACKNOWLEDGMENTS

I thank Valerie Hans, James Levine, Leslie Ellis, Jason Schklar, two anonymous reviewers, Ron Roesch, and Bette Bottoms for their comments and suggestions, and my collaborators Jonathan Casper, Stephan Landsman, Michael Saks, and Linda Dimitropoulos, who stimulated my evolving thoughts about jury simulations.

### REFERENCES

- Arizona Supreme Court Committee on More Effective Use of Juries (1994). *Jurors: The power of twelve*. Phoenix: Arizona Supreme Court Committee.
- Baumeister, R. F., & Darley, J. M. (1982). Reducing the biasing effect of perpetrator attractiveness in jury simulation. *Personality and Social Psychology Bulletin*, 8, 286-292.
- Brekke, N. J., Enko, P. J., Clavet, G., & Sellau, E. (1991). Of juries and court-appointed experts: The impact of nonadversarial versus adversarial expert testimony. *Law and Human Behavior*, 15, 451-475.
- Casper, J. D., Benedict, K., & Perry, J. L. (1989). Juror decision-making, attitudes, and the hindsight bias. *Law and Human Behavior*, 13, 291-310.
- Connolly, T., & Bukszar, E. W. (1990). Hindsight bias: Self flattery or cognitive error? *Journal of Behavioral Decision Making*, 3, 205-211.
- Cowan, C. L., Thompson, W. C., & Ellsworth, P. C. (1984). The effects of death qualification on jurors' predisposition to convict and on the quality of deliberation. *Law and Human Behavior*, 8, 53-79.
- Daubert v. Merrell Dow Pharmaceuticals, 113 S.Ct. 2786 (1993).
- Diamond, S. S. (Ed.). (1979). Simulation research and the law. A special issue of *Law and Human Behavior*, 3.
- Diamond, S. S., & Casper, J. D. (1992). Blindfolding the jury to verdict consequences: Damages, experts, and the civil jury. *Law & Society Review*, 26, 513-563.
- Diamond, S. S., Casper, J. D., Heiert, C., & Marshall, A. M. (1996). Juror reactions to attorneys at trial. *Journal of Criminal Law and Criminology*, 87, 17-47.
- Diamond, S. S., & Levi, J. N. (1996). Improving decisions on death by revising and testing jury instructions. *Judicature*, 79, 224-232.
- Ellsworth, P. C. (1988). Unpleasant facts: The Supreme Court's response to empirical research on capital punishment. In K. C. Haas & J. A. Inciardi (Eds.), *Challenging capital punishment: Legal and social science approaches*. (pp. 177-211). Newbury Park, CA: Sage.
- Ellsworth, P. C. (1989). Are twelve heads better than one? *Law and Contemporary Problems*, 52, 205-224.
- Elwork, A., Sales, B. D., & Alfini, J. (1982). *Making jury instructions understandable*. Charlottesville, VA: Michie.

- Federal Judicial Center (1994). *Reference manual on scientific evidence*, Washington, DC: Federal Judicial Center.
- Fischhoff, B. (1975). Hindsight  $\neq$  foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.
- Free v. Peters, 806 F. Supp. 705 (N. D. Ill. 1992); *rev'd*, 12 F.3d 700 (7th Cir. 1993), *cert. denied*, 513 US 967 (1994).
- Goldberg, F. (1970). Toward expansion of *Witherspoon*: Capital scruples, jury bias, and the use of psychological data to raise presumptions in the law. *Harvard Civil Rights-Civil Liberties Review*, 5, 53-69.
- Hamilton, D., & Sherman, J. S. (1996). Stereotypes. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed.) (pp. 1-68). Hillsdale, NJ: Erlbaum.
- Hans, V. P., & Lofquist, W. S. (1992). Jurors' judgments of business liability in tort cases: Implications for the litigation explosion debate. *Law & Society Review*, 26, 101-115.
- Hastie, R., Penrod, S. D., & Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Harvard University Press.
- Heuer, L., & Penrod, S. (1994). Juror notetaking and question asking during trials: A national field experiment. *Law and Human Behavior*, 18, 121-150.
- Kalven, H., Jr., & Zeisel, H. (1966). *The American jury*. Boston: Little, Brown.
- Kaplan, M. F., & Krupa, S. (1986). Severe penalties under the control of others can reduce guilt verdicts. *Law and Psychology Review*, 10, 1-18.
- Kaplan, M. F., & Miller, L. E. (1978). Reducing the effects of juror bias. *Journal of Personality and Social Psychology*, 36, 1443-1455.
- Kassin, S. M., & Wrightsman, L. S. (1979). On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology*, 37, 1877-1887.
- Kerwin, J., & Schaffer, D. R. (1994). Mock jurors versus juries: The role of deliberations in reactions to inadmissible testimony. *Personality and Social Psychology Bulletin*, 20, 153-162.
- Kramer, G. P., & Kerr, N. L. (1989). Laboratory simulation and bias in the study of juror behavior: A methodological note. *Law and Human Behavior*, 13, 89-99.
- Linz, D., & Penrod, S. (1982). *A meta-analysis of the influence of research methodology on the outcomes of jury simulation studies*. Paper presented at the Academy of Criminal Justice Sciences, Louisville, Kentucky.
- Lockhart v. McCree, 476 U.S. 162 (1986).
- MacCoun, R. J., & Kerr, N. L. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology*, 54, 21-33.
- McGlynn, R. P., Megas, J. C., & Benson, D. H. (1976). Sex and race as factors affecting the attribution of insanity in a murder trial. *Journal of Psychology*, 93, 93-99.
- Mitchell v. Gonzales, 819 P.2d 872 (Cal. 1991).
- People v. Allen, 420 N.W.2d 499 (Mich. 1988).
- Sandys, M., & Dillehay, R. C. (1995). First-ballot votes, pre-deliberation dispositions, and final verdicts in jury trials. *Law and Human Behavior*, 19, 175-195.
- Sears, D. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515-530.
- Severance, L. J., Greene, E., & Loftus, E. F. (1984). Toward criminal jury instructions that jurors can understand. *Journal of Criminal Law and Criminology*, 75, 198-233.
- Stasser, G., Kerr, N., & Bray, R. M. (1982). The social psychology of jury deliberations: Structure, process, and product. In N. Kerr & R. M. Bray (Eds.), *The Psychology of the courtroom* (pp. 221-256). New York: Academic Press.
- Ugwuegbu, D. C. E. (1979). Racial and evidential factors in juror attribution of legal responsibility. *Journal of Experimental Social Psychology*, 15, 133-146.
- Vidmar, N. (1979). The other issues in jury simulation research: A commentary with particular reference to defendant character studies. *Law and Human Behavior*, 3, 95-106.
- Vidmar, N. (1994). Making inferences about jury behavior from jury verdict statistics: Cautions about Lorelei's Lied. *Law and Human Behavior*, 18, 599-617.
- Walker, L., & Monahan, J. (1996). *Daubert and the Reference Manual: An essay on the future of science in law*. *Virginia Law Review*, 82, 837-857.
- Weiten, W., & Diamond, S. S. (1979). A critical review of the jury simulation paradigm: The case of defendant characteristics. *Law and Human Behavior*, 3, 71-93.
- Witherspoon v. Illinois, 391 U.S. 510 (1968).
- Zeisel, H. (1968). *Some data on juror attitudes toward capital punishment*, Chicago: Center for Studies in Criminal Justice, University of Chicago Law School.