

## **Sticks and Stones**

**Phoebe C. Ellsworth<sup>1</sup>**

I believe that research should be refuted by research. More and more of our scarce journal space is being taken up by attacks, rebuttals, and rebuttals to the rebuttals, often ending with a whimper of recognition that the adversaries were not so very far apart to begin with, and that the only way (if possible) to resolve the disagreement is through empirical research. Communication of scientific disagreement does not require a published article. Grant proposals and manuscripts submitted to refereed journals like this one are sent out to reviewers, who provide written evaluations that are communicated to the author. Papers presented at conferences are evaluated and often criticized by discussants and members of the audience. The best strategy, I think, is direct communication between the critic and the author. Unlike the published criticism, this method allows both critic and author to reach a clear understanding of each other's position, to identify specific topics of agreement and disagreement, and perhaps even to arrive at possible resolutions through the normal give and take of scientific dialogue. The best a published criticism can do is to provide an opportunity for back-to-back monologues, surely a far less effective means of communication.

I also feel that the attack-and-rebuttal format gives too much of an advantage to the critic at the expense of the original author, who, after all, is the one who put in all the hard work of carrying out the research in the first place. The critic gets to define the time line, taking as long as she or he wants to work on the critique, consulting with other like-minded scholars, and sending it off when it is ready. The author is often completely unaware that any such critique is in the works until it arrives in the mailbox with a letter from the editor graciously offering the author a chance to reply to the criticism, and asking that the reply be sent back within 3 weeks or a month. Very few of us are sitting around wondering how we are going to occupy ourselves for the next month. Students are waiting for grades or recommendations or feedback on their dissertations, committees are faced with urgent decisions, families demand attention, and we are already past the deadline on several other commitments. The author is forced to drop everything and respond immediately, or forego the opportunity to respond, leaving the readers to conclude

<sup>1</sup>Psychology Department and Law School, University of Michigan, Ann Arbor, Michigan 48109; e-mail: pce@umich.edu

that the critique must be devastatingly accurate, since the author could come up with no defense.

The Adversarial Forum is a useful forum for the expression of opinions about general issues of theory, method, application, and policy. The *general* controversy over the so-called "litigation crisis" and the real or imagined escalation in damage awards would be a fine topic for an Adversarial Forum. I would not like to see the Forum turn into a platform for attacks on individual research studies.

*Every* empirical study has flaws. Almost every author exaggerates the significance of the problem studied, overstates the implications of the results, or both. The Hastie, Schkade, and Payne study is no exception. As authors we make unsubstantiated claims, as readers we recognize them, and are unlikely to be fooled. Vidmar is quite right that we cannot be confident that the conclusions drawn by Hastie and his colleagues are true, but of course his critique provides no evidence that they are not.

But perhaps the Hastie, Schkade, and Payne study had particularly serious weaknesses that somehow slipped by the editor and the reviewers and that would not be apparent to readers. Perhaps it was not up to the usual standards of *Law and Human Behavior*. If this were the case, some people might believe that a published criticism was warranted, although I would not be one of them.

There are many strengths to the study. The participants were not college students, but an unusually representative sample of 726 jury-eligible citizens. To enhance generalizability, four different stimulus cases were used, all based on real cases, two of which were originally decided by juries. The researchers examined both the verdicts of individual jurors and the verdicts of deliberating juries, looking for correlates of both. The jury deliberations were videotaped, transcribed, and carefully analyzed in an attempt to understand the individual and group processes by which juries reached their decisions. In fact, the methodology of this study is superior to that of most published jury studies. If this study warrants a published criticism, then so do dozens of others, and we run the risk of turning *Law and Human Behavior* into a journal with as many critiques as reports of original research.

There are also weaknesses in the study, although as usual the reviewers disagreed about which were the most important. For what it's worth, in my opinion the most serious problem is that the jurors were not given the chance to decide compensatory damages, only punitive damages. Making both decisions, as real juries do, may make the distinctions between compensatory and punitive damages more noticeable, thereby improving accuracy of recall. In addition, the opportunity to make substantial compensatory damage awards might have restrained jurors' motivation to make large punitive damage awards. In this study, the information about the prior compensatory damage awards was not very salient; for example, participants were not told how much money the plaintiffs had already been awarded. Whether this change in procedure would actually make a difference is of course an empirical question, one that I hope Hastie will address in follow-up research.

The 5% accuracy rate for memory for the law also raised questions in my mind. Jury researchers are well aware that jurors' inability to understand the law is their greatest weakness, but 5% accuracy is conspicuously worse than even the usual poor performance level, and requires some explanation. Were the questions unusually

difficult, the scoring particularly stringent, or what? Is there reason to believe that this reflects an actual substantive difference, or might it be methodological? This startling number deserved more attention in the original write-up, and deserves more careful exploration in follow-up research.

As for Vidmar's central criticism—that real juries would never be asked to make these decisions in the first place—the problem is that in two of these cases they *were* asked to make these decisions. Maybe the fault is in the gatekeeping function the judges are supposed to exercise, and maybe the juries did badly because they never should have been asked to make decisions like these, but whether the problem is that juries are reversed because the task is too hard for them or because they are given inappropriate cases does not have much *practical* significance. Its theoretical significance depends on whether juries do well in appropriate cases and poorly in inappropriate ones. Neither the original study nor the critique sheds any light on this question.

Would it have been better to include cases where the courts held that punitive damages *were* appropriate? Sure. If the juries decided that the damages were not appropriate then it would look as though juries may be *generally* incompetent to decide punitive damages as the task is currently structured. If the juries agreed that damages were appropriate, then it would look as though juries may have a pro-plaintiff bias, deciding for the plaintiff whether or not that was the right decision. Such a study could illuminate the specific nature of the jury's incompetence; it could not provide evidence for jury competence. If this additional question were addressed in a balanced design it would require a total of 1,452 participants, rather more than we typically require in a single journal article. Just as every study has flaws, every study necessarily leaves some questions unanswered.

And of course Hastie did not design this study to answer all possible questions about the way juries decide punitive damages. Hastie, Schkade, and Payne is just one study in a program of research. Different cases are used in other studies in the series, and different questions are asked. The basic idea of systematic replication and extension is to provide convergent validity, using a variety of designs and procedures that compensate for each other's weaknesses, relying on later studies to address the questions left unanswered in the earlier ones, and testing the generality of the results with different subject populations and stimulus materials. When an author has conducted several studies on the same topic, the critic has an obligation to consider the whole body of research, rather than singling out just one study. The reader of Vidmar's critique would never know that there *are* other studies in the series. Perhaps Vidmar himself did not know, but that just reinforces my original point that dialogue is a far more useful method than published critiques for resolving scientific disputes.

Finally, I want to take issue with Vidmar's statement that "judgment following verdict, appellate review, and private settlement between the parties provide effective and legal and pragmatic constraints on wayward jury verdicts" (p. 705). I do not trust the argument that discovery and correction of errors in initial decisions is evidence that the system is working as it should be. Undoubtedly this mistrust stems from my work on capital punishment. Since executions resumed in the late 1970s, dozens of people have been released after years on death row because the

initial decision that they were guilty turned out to be a mistake. Prosecutors make pious claims that the discovery of the mistake while the defendant was still alive is evidence that “the system is working.” But mistakes are extremely costly even when they are not deadly. Much more important, we have no way of knowing how many mistakes are never corrected.

#### ACKNOWLEDGMENTS

This article is an elaboration of my review of Vidmar’s original manuscript, and, since I have also been a reviewer of several of the studies coming from Hastie’s program of research on punitive damages, the editor asked if he could published my reactions. Of course my basic argument is that this is exactly the sort of thing I think should *not* take up journal space, so I am as guilty as the next person. I wasted a lot of time wrestling with this inconsistency, and Barbara Zezulka Brown still managed to help me get the manuscript done on time.

Support for this article was provided by the Cook Fund of the University of Michigan Law School.