

Michigan Law Review

Volume 79 | Issue 4

1981

Judgment *Non Obstantibus Datis*

Reid Hastie
Northwestern University

Follow this and additional works at: <https://repository.law.umich.edu/mlr>



Part of the [Comparative and Foreign Law Commons](#), [Courts Commons](#), and the [Law and Psychology Commons](#)

Recommended Citation

Reid Hastie, *Judgment Non Obstantibus Datis*, 79 MICH. L. REV. 728 (1981).
Available at: <https://repository.law.umich.edu/mlr/vol79/iss4/13>

This Review is brought to you for free and open access by the Michigan Law Review at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

JUDGMENT *NON OBSTANTIBUS DATIS*

*Reid Hastie**

JURY TRIALS. By *John Baldwin* and *Michael McConville*. Oxford: Oxford University Press. 1979. Pp. vii, 150. \$12.95.

Jury Trials is a flawed book that reveals many of the obstacles confronting social scientists who attempt to resolve legal issues through empirical research. The goal of the research reported by Baldwin and McConville is "to evaluate the performance of the jury in a series of criminal trials" (p. 21). More specifically, it is "to pit, as it were, the verdict of the jury against the verdict of others involved in the case" (p. 21). Within this comparative framework, three empirical questions were posed: (1) What are the evidentiary or case circumstances in which other participants in the trial are likely to express high rates of disagreement with jury verdicts? (2) To what extent can questionable verdicts (verdicts questioned by other trial participants) be explained with reference to the composition of the jury? (3) Are professional criminals more likely than others to avoid conviction, and are they frequently acquitted in questionable jury verdicts?

I

The method Baldwin and McConville selected to answer their questions was straightforward. Evaluations of jury verdicts in approximately 700 criminal trials (370 in Birmingham and 358 in London) were obtained from trial judges, solicitors, defendants, and police officers who were involved in the cases. The respondents were not all equally eager to participate in the research, and response rates and interview formats varied considerably across groups of respondents. For example, the London sample included only police officers, while in Birmingham a small percentage of defendants and defense solicitors responded (barristers refused to participate at all); police and defendant data were obtained in face-to-face interviews, while solicitors and judges responded only on written questionnaires.

Because of constraints placed on the research by legal authorities, Baldwin and McConville could not ask direct questions about the

* Associate Professor of Psychology, Northwestern University. B.A. 1968, Stanford University; Ph.D. 1973, Yale University. — Ed.

respondents' evaluation of the verdict. Instead, the researchers distributed a written questionnaire that included a checklist of factors that might explain the jury verdict, and asked the respondent to indicate which factor or factors explained the verdict in a particular trial. For example, the possible explanatory factors for a guilty verdict included: strength of the prosecution case, weakness of the defense case, absence of defense witnesses, failure of the defense to call witnesses, unreliability of defense witnesses, and bad impression created by the defendant (p. 28). A final item asked the respondent to comment on the case. The report does not describe the face-to-face interview schedule, but it is clear that interview questions were more direct. In addition to the basic verdict evaluation data, the researchers collected information about jury composition (each juror's age, sex, occupation, race, and number of previous jury cases), case outcomes, case preparation, and preverdict predictions from some solicitors. However, the book provides detailed summaries of only the verdict evaluation and jury composition data.

It is obvious that this method sharply limits conclusions from the research. The most important limits are created by the biased sampling plan, varying response rates, and differences in the amount and types of information obtained from the subgroups in the sample. For example, we would expect that police views would dominate the results, that the defense point of view would be underrepresented, and so forth. In addition, the written questionnaire can provide weak support, at best, for inferences about the quality of jury verdicts. The authors themselves were discomfited by the indirectness of their questions; we might magnify their discomfort by noting that laboratory research has shown that individuals' reports of their causal inference processes appear to be of uncertain validity.¹ If it seems that *self*-reports are invalid, what are we to make of a solicitor, judge, or police officer's insights into a juror's reasoning processes?

A further concern is raised by the authors' use of the questionnaire responses (mostly in the form of checkmarks) to classify jury verdicts as questionable or dubious. The questionnaires were not designed to identify improper verdicts and the authors do not describe the classification rules that they applied to infer them. Because of this omission, we never learn how the major distinction in the analysis (the proper versus the improper verdict) was defined.

1. See, e.g., Nisbett & Wilson, *Telling More Than We Can Know: Verbal Reports On Mental Processes*, 84 PSYCH. REV. 231, 242 (1977).

The researchers' analysis began with the identification of "questionable" or "doubtful" verdicts followed by a search for factors to explain these "miscarriages of justice." In the Birmingham sample, "questionable acquittals" were marked whenever the trial judge and at least one other respondent indicated that an acquittal was not justified (pp. 46-47). Forty-one of the acquittals (thirty-six percent) were labeled questionable under this procedure. A verdict was classified as a "doubtful conviction" when at least two respondents (not necessarily including the judge) doubted the jury's verdict. Fifteen of the convictions (six percent) were so classified. The researchers do not explain how they determined that a respondent believed the verdict was incorrect, and the answer is not obvious given the indirect nature of the questioning. Interestingly, if the authors had applied the same standards for convictions as for acquittals, a maximum of eight convictions would have been classified doubtful (p. 71). The method used to identify questionable acquittals (twenty-five percent, thirty-nine verdicts) and doubtful convictions (six percent, ten verdicts) in the London sample, where police officers were the only respondents, is even more obscure.

The authors' major conclusion from the verdict disagreement results is that "juries in both Birmingham and London were thought by respondents to have reached wrong or questionable conclusions on the evidence with a surprising frequency" (p. 127). The authors' efforts to account for the appearance of "questionable acquittals" were fruitless:

[We] could not, for example, explain the questionable outcomes in terms of the social composition of the jury; we could see little evidence of jury equity or of juries disliking particular laws; it did not seem that professional criminals were often able to secure wrongful acquittals. In short, almost no overall patterns were evident and no common factor or factors appeared to explain the questionable outcomes we encountered. [P. 131.]

As to doubtful convictions, the authors felt that their statistic indicating that "at least five percent of those convicted by jury were said to have been convicted in doubtful circumstances" was the "most disturbing finding to emerge from [the] research" (p. 128). They concluded that two factors accounted for the occurrence of "doubtful convictions":

[F]irst, that the jury appeared in many of these cases to be too easily satisfied of the defendants' guilt because they failed to appreciate the high standard of proof required in criminal cases; second, that the jury apparently convicted the defendant through lack of comprehension of the issues involved. [P. 76.]

The support for this two-factor explanation is summarized as a series of remarks made by the (nonjuror) respondents, and it is weak. It is not clear how well the selection of remarks represents all respondents' views; many of the remarks cited do not strongly implicate the authors' favored factors, and, at best, the remarks are speculations generated by officials who were not involved in the jury's deliberations. For example, Baldwin and McConville quote a police officer's remark that "I wouldn't have been surprised if they had acquitted him. His story was more than plausible and he was very good in evidence. Our case was quite weak and I was surprised at the verdict. I have been generally satisfied with the jury, but not on this occasion" (p. 78). Three questions immediately come to mind: Does this quotation accurately represent all this officer's reactions to this verdict? Does it indicate that the officer felt that the jury "failed to appreciate the high standard of proof required in criminal cases," or might it mean that he believed some other factor explained the "doubtful conviction?" How would a police officer have an accurate, complete view of the trial evidence or of the jury's decision processes? These same questions can be raised for all of the anecdotes cited in support of the two-factor explanation, and the authors do not answer them.

The authors cite a third factor, racial prejudice, that might have influenced jury deliberations. The defendant was black in seventeen percent of the questionable acquittals (seven of the forty-one) and in fifty-three percent of the doubtful convictions (eight of fifteen) (p. 81). A chi-squared test for independence of frequencies in a race (black-nonblack) by verdict type (questionable acquittal-doubtful conviction) contingency table yields a value of 8.36, significant at the $p < .01$ level. The authors did not calculate this statistic. They note that, "Given the small number of cases involved, however, even this pattern must be regarded as tentative in nature" (p. 131).

I have noted that the authors do not clearly specify the methods used to identify questionable or doubtful verdicts, and they do not use comparable standards for classifying acquittals and convictions as in error. There are at least four other serious methodological problems with using Baldwin and McConville's data to detect relationships between questionable or doubtful verdicts and possible explanatory factors. First, the search for explanatory factors focuses only on the improper verdict cases. The analysis is not comparative; features of the cases classified as questionable or improper are cited, but we are given no statistical or impressionistic comparison to features of the proper-verdict cases. Thus, for example, one cannot re-

late the statistics for improper conviction and acquittal rates of black defendants to the central conceptual question (What causes improper verdicts?) because base-rate statistics on the fate of black defendants in proper verdict cases are not supplied: an interpretation of the fifty-three percent doubtful conviction rate would vary considerably if the base-rate for proper convictions were ten, fifty, or ninety percent. Second, the small sample size of fifty-six cases lowers the internal power of analyses within the sample and the external power for generalization to unsampled cases. A less obvious third problem is that the search for explanatory factors is not systematic. For example, we do not know how many candidate-explanatory factors were considered in the analyses of potential case-dependent factors (chapters 4 & 5) or jury factors (chapter 6). Thus, when an apparent correlation between race and type of verdict error (questionable acquittal, doubtful conviction) appears, we cannot estimate its significance because we do not know how much multiple testing has inflated the chances of a spurious significant result. Finally, the authors could look only to five background variables (juror sex, age, occupation, race, and number of previous impanelments) to analyze the effects of jury composition on verdicts. They had no other background data, and no information concerning individual behavior during deliberations. Given this meager information and the methodological problems described above, it is not surprising that the authors failed to find relations between verdict and jury composition.

II

Up to this point this Review has been a catalog of methodological weaknesses that reads like the table of contents for a research methods text: biased sampling plan, unequal response rates, defective questionnaire design, obscure and uneven operational classification rules, impressionistic rather than quantitative analysis, focus on the positive cell to answer a correlational question, failure to report relevant base-rate data, inadequate sample size, unsystematic consideration of multiple predictive factors, and use of aggregated data to answer questions about individual behavior. Although the authors acknowledge many of these problems, they do not accept the strong implications that their conclusions rest on impressions rather than rigorous analysis, that many subtle and distorting biases are present in their methods, and that their failure to discover clear empirical relationships was predetermined by the research plan. Instead they conclude:

The evidence that we have presented in this book has shaken our own

confidence in the system of trial by jury. We believe that our own inquiry has raised a sufficient number of important questions about the precision of trial by jury to justify giving researchers access to the jury room. . . . The onus to demonstrate why this should not be done would now appear to lie with those who argue for its continued immunity from research. [P. 132.]

If the authors recognize that their research is riddled with many methodological flaws, why do they present their conclusions with such confidence? The answer to this question is the critical lesson to be learned from *Jury Trials*. The answer is important because the authors have presented strong conclusions with great significance for legal policy. The methodologically unsophisticated reader may be tempted to accept the authors' conclusion that the jury trial is a defective, error-prone decision mechanism because the authors speak with the authority of social scientists. I want to make it clear that I object not to this conclusion but to the spurious use of scientific authority to support the conclusion. The jury *may or may not* be an effective dispute resolution mechanism, but the authors' research results are definitely not an empirical, scientific demonstration that jury verdicts are error-prone. I think that such a misuse of research results may be the greatest sin that can be committed by a social scientist conducting problem-oriented research. Why does it happen?

Several factors underlie the authors' spurious sense of confidence in the results of their research, and these same factors may inflate confidence in the results of any research. First there is the simple political and economic reality of the applied researcher. Social scientists often make their living by planning and executing research projects. To obtain and maintain funding they must express an air of confidence in their plans, their methods, and their conclusions. It is virtually fatal to display self-doubt, to abort a plan once funding is started (regardless of the limits placed on their research instruments, the lack of cooperation from participants, etc.), or to bury a final report even when it is clear the method is intolerably weak. These researchers have to report strong conclusions or they will be out of work.

Second, pervasive and subtle biases affect the judgments of both naïve and sophisticated scientists.² Human judgments are dominated by recent, salient, frequent, memorable, and emotionally

2. See generally R. HOGARTH, JUDGMENT AND CHOICE 155 (1980); R. NISBETT & L. ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS ON SOCIAL JUDGMENT 249 (1980).

arousing sources of information.³ Overconfidence in one's judgment is an almost irresistible consequence of decision making without feedback.⁴ Experience — even training in statistics and research methods — does not eliminate judgmental biases resulting from preconceptions, sampling error, and systematic measurement error.⁵ The authors' overdependence on police officers' views, blindness to the importance of base-rate data, and overinterpretation of verbal comments are common examples of judgmental biases, biases that elude conscious detection.

Third, the authors started the research program with an impossible question. To rephrase the logic of their investigation, they hoped to identify cases in which the jury verdict was in error and then to explain these errors with reference to characteristics of the jury. It is not clear that any empirical research method could adequately meet these goals, but the inadequacy of this method is underscored by the fact that jurors were not even included among the respondents. Perhaps Baldwin and McConville's research could have provided insights into factors that affect the attitudes of the judge, the solicitor, the defendant, and (most of all) the police officer toward a trial verdict. The authors might have proposed a theory to explain the sizeable discrepancies among these perspectives; their interview data might have been relevant to such a theory. But, their belief that interviews with these trial participants provide valid evidence concerning jury behavior is simply unacceptable.

What are the implications of these factors for the design of future research on jury behavior? The first strong implication is that policymakers and others who fund projects such as *Jury Trials* should become more sensitive to the realities of research. How to educate policymakers is unclear, but the consequence of ignorance is not: it will result in more research like that reported in *Jury Trials*.

The second implication is that researchers should be aware of their own judgmental biases and attempt to correct for them when drawing conclusions from data. This is a difficult task. It is decep-

3. Taylor & Fiske, *Saliency, attention, and attribution: Top of the head phenomena*, in 11 *ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY* 253 (L. Berkowitz ed. 1978); Tversky & Kahneman, *Availability: A heuristic for judging frequency and probability*, 5 *COGNITIVE PSYCH.* 207, 230 (1973).

4. See, e.g. Einhorn & Hogarth, *Confidence in Judgment: Persistence of the Illusion of Validity*, 85 *PSYCH. REV.* 395, 409 (1978).

5. See, e.g., Tversky & Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 *SCI.* 1124, 1130 (1974), and Tversky & Kahneman, *Belief in the Law of Small Numbers*, 76 *PSYCH. BULL.* 105, 109 (1971).

tively easy to acknowledge the existence of a general bias but perversely difficult to identify its effects in one's own thought processes. Certainly it helps to be aware of the existence of general classes of bias. In addition, the relentless application of statistical procedures as well as the intuitive evaluation of conclusions is a strong safeguard.

A third implication is that the initial step in the research program — statement of research questions — is of utmost importance. Two fine essays illustrate this point. In his essay on the "Quest for the Middle Range"⁶ Harry Kalven, Jr. clearly and forcefully argued that the social scientist should avoid research questions that concern high-level premises about fundamental values and preferences because those issues are beyond the reach of empirical facts. Similarly, he argued that certain premises are beneath empirical evaluation because they are trivially correct and require no additional confirmation. Instead, the researcher should frame research questions in the middle range, where empirical results can sharpen controversy by eliminating some disagreements, by revealing new conceptual relations, and by focusing debate on nonempirical issues. Baldwin and McConville's research program that attempts to explain inaccurate jury verdicts by studying nonjuror reactions to the verdicts surely falls outside this middle range. It is probable that no adequate empirical criterion for verdict accuracy could be established; it is certain that none was developed in *Jury Trials*. A second wise comment on research questions comes from a recent essay by Phoebe Ellsworth.⁷ Ellsworth argues that one must consider not only the form of the question addressed by research, but also the appropriateness of the research method employed to answer the question. Thus, perfectly good questions and perfectly good methods may still result in useless research programs when inappropriately matched. In the present example, I would guess that even an acceptably restated version of the accuracy question could not be addressed by Baldwin and McConville's indirect interview method. Post-trial interviews with jurors or a mock-trial would be more appropriate approaches.

Kalven's classic essay included another message of particular relevance to books like *Jury Trials*. Empirical researchers should not expect to resolve legal policy questions solely or even directly with

6. See generally Kalven, *The Quest for the Middle Range: Empirical Inquiry and Legal Policy*, in *LAW IN A CHANGING AMERICA* (G. Hazard ed. 1968).

7. Ellsworth, *From Abstract Ideas to Concrete Instances: Some Guidelines for Choosing Natural Research Settings*, 32 *AM. PSYCH.* 304, 309 (1977).

empirical facts. Legal policy disputes inevitably involve a multiplicity of "legal ends"; their resolution invariably requires a consensus concerning preferences. Empirical truths will, only sometimes, be relevant.