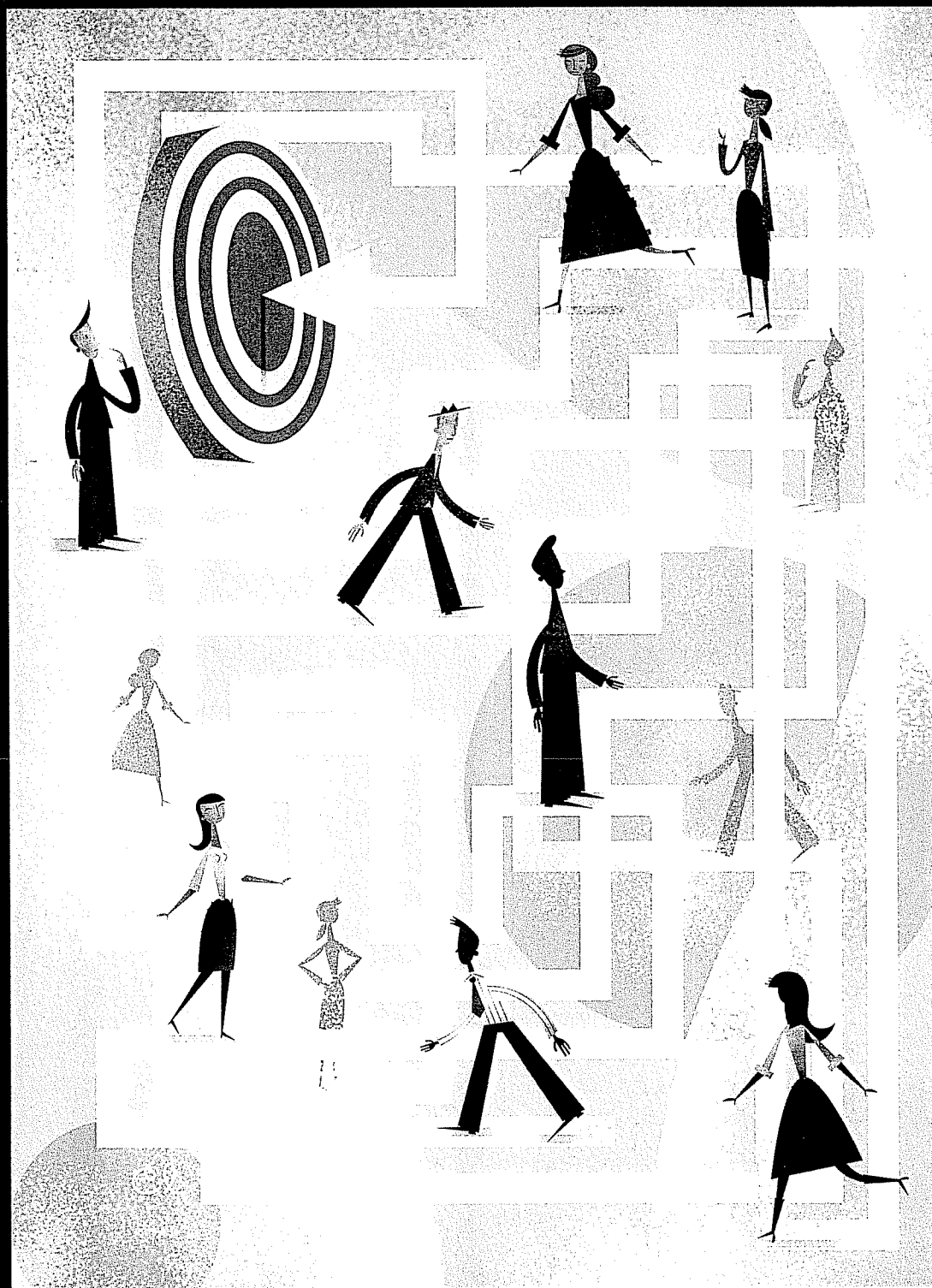


VOIR DIRE

A PUBLICATION OF THE AMERICAN BOARD OF TRIAL ADVOCATES



COVER STORY:

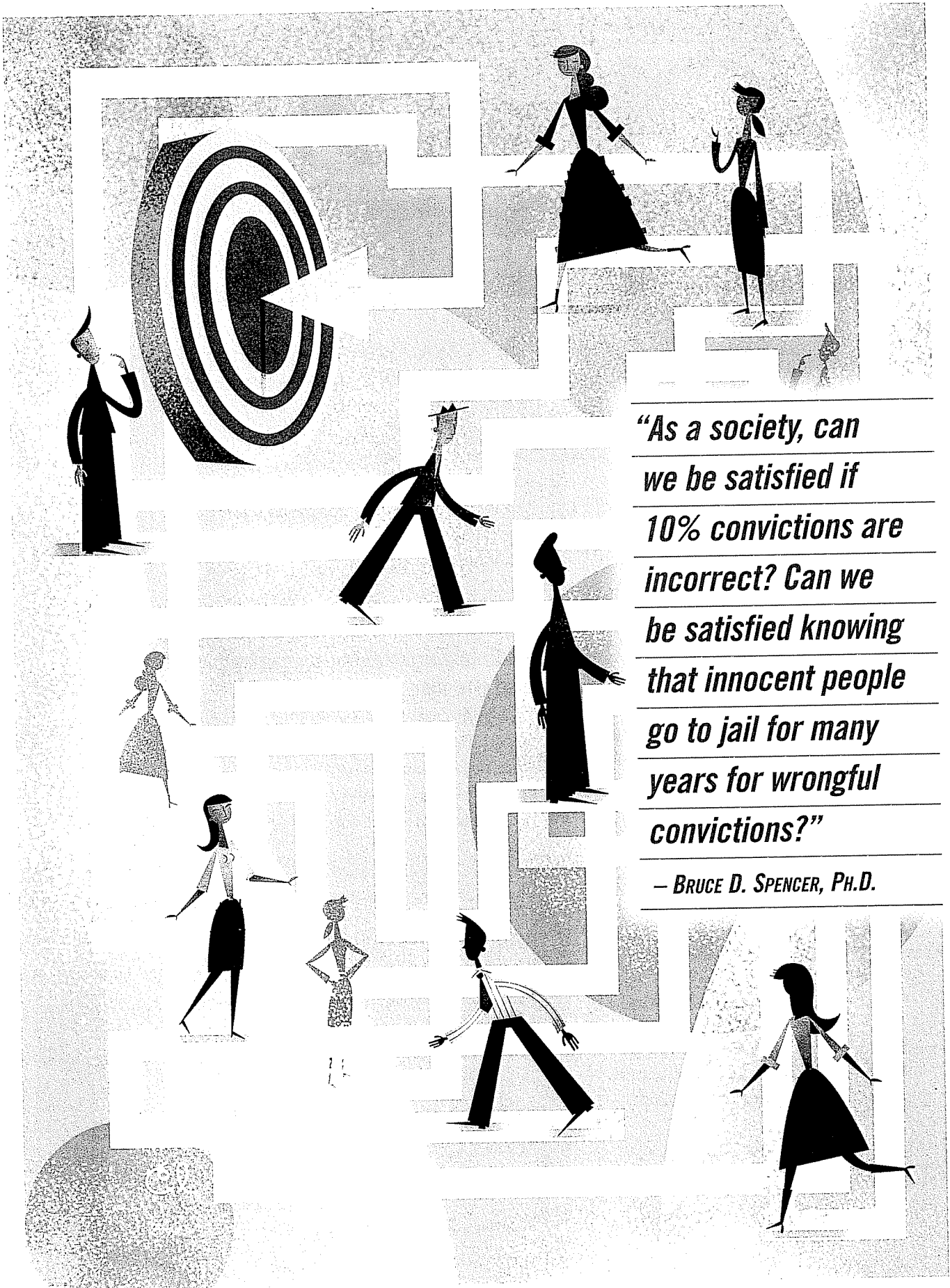
On Second Guessing Jury Verdicts

PLUS

*Jury Verdicts
and the Need
for Tort
Reform*

*More on the
McDonald's
Coffee
Lawsuit*

*Judicial
Independence
and Public
Education*



“As a society, can we be satisfied if 10% convictions are incorrect? Can we be satisfied knowing that innocent people go to jail for many years for wrongful convictions?”

— BRUCE D. SPENCER, PH.D.

By Paula Hannaford-Agor^{1*}

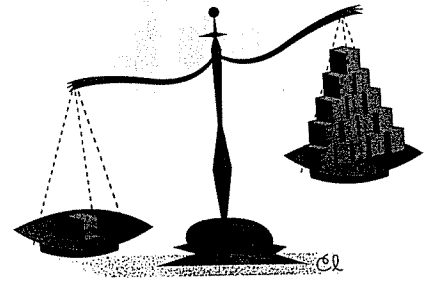
America has something of a love-hate relationship with juries. We are at once effusive in our praise of the institution of trial by jury, yet highly suspicious of the integrity and intelligence of individual jurors and the accuracy of jury verdicts in general. The American legal system has extraordinary confidence in the fact-finding ability of juries and is highly resistant to overturning jury verdicts. Yet it is only to be expected that any system that relies exclusively on human judgment will make a

non-random sample of 290 felony jury trials that took place in four large, urban courts in 2000-2001, Spencer estimates that 87% of jury verdicts are correct, which he implies is an unacceptably low rate.³ Moreover, he concludes that when juries err, they do so asymmetrically; they are more likely to convict an innocent person than to acquit a guilty person.⁴

All in all, Spencer's article is a masterful application of statistical modeling that appears, at long last, to quantify the proportion of juries' errors with far greater precision than the appellate bench could ever dream. Unfortunately, the model itself has

hung juries. His models make use of the judge-jury agreement rate for 290 non-capital felony trials that took place in four large, urban courts in 2000-2001,⁶ as well as that for the more than 3,500 criminal trials included in Kalven and Zeisel's classic study of the American jury, published in 1966.⁷ Detailed information about the NCSC trials, including judges' reports about what their decisions would have been had the cases been tried as bench trials, was collected through questionnaires distributed to judges, lawyers, and jurors in a study designed to examine the frequency and causes of hung juries.⁸ The

ON SECOND GUESSING JURY VERDICTS



mistake now and then. The possibility of an "incorrect verdict" is troubling to policy makers concerned with the integrity of the justice system – and rightly so as there are enormous consequences – not only for the "losing" party, regardless of who that might be, but also for the greater community. The number of convicted felons who have been exonerated by improved DNA and other forensic tests, sometimes after years in prison or even on death row, is a sobering reflection on the imperfections of our justice system.

Litigants, their lawyers, sometimes large numbers of casual observers in the community, and even the trial judge (privately) might disagree with the jury's decision in a given case, but Professor Bruce D. Spencer, a Faculty Fellow of the Northwestern University Institute for Policy Research, has recently attempted to determine just how often juries "get it wrong" using sophisticated statistical modeling techniques.² Based on a

serious conceptual and methodological problems that raise serious questions about Spencer's findings and conclusions about jury accuracy. After describing Spencer's study in greater detail, I will focus most of my attention on the conceptual problems, which I believe are the more serious. Then I will turn to some methodological problems and the overall implications of Spencer's study for the American justice system.

A Statistical Study of Jury Accuracy

In his article "Estimating the Accuracy of Jury Verdicts," Spencer employs statistical modeling techniques to calculate the probability that a jury will either convict an innocent person or acquit a guilty person, even if the actual guilt or innocence of the person is unknown.⁵ Spencer's analyses are based on data collected by the National Center for State Courts (NCSC) for a study of

Kalven and Zeisel study collected information about judges' preferences for trial outcomes, but did not collect information about how jurors reached their decisions. The NCSC study also collected information about judges', lawyers', and jurors' assessments about the strength of the evidence and the credibility of trial witnesses.

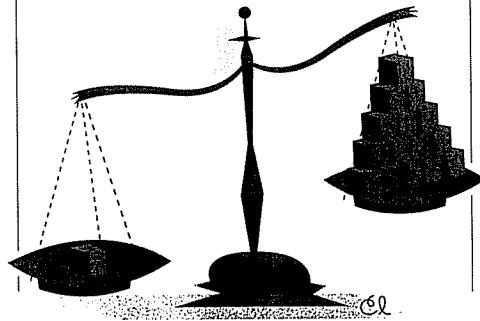
Spencer argues that it is possible to determine the accuracy of jury verdicts – that is, the proportion of verdicts that are incorrect – from the judge-jury agreement rate.⁹ The logic on which the statistical models are premised is fairly straight-forward. If a judge and jury, after observing the same trial and considering the same evidence, disagree on what the verdict should be, one of them must be right and the other wrong. From a procedural viewpoint, a "correct" acquittal results only if the prosecution fails to prove beyond a reasonable doubt that the defendant committed the crime (even if he, in fact, did so). From an "omniscient" perspective,

however, a verdict to convict is correct if the defendant in reality committed the crime, even if the prosecution presented no evidence at trial. The same logic would apply in a civil trial: a verdict for an injured plaintiff could be defined as correct if the plaintiff proved by a preponderance of the evidence all the elements of his claim or, alternatively, if the defendant negligently caused the plaintiff's injury regardless of the evidence presented by the plaintiff at trial. According to Spencer, how one defines a "correct verdict" does not affect the estimates of jury accuracy.¹⁰

Although Spencer entitles his article "Estimating the Accuracy of Jury Verdicts," he does acknowledge that the judge's preferred verdict isn't necessarily the correct one. Judges and juries, after all, bring different types of expertise and perspectives to their tasks. Judges are not only legal experts, but also have the advantage of becoming extremely knowledgeable about evidence that is presented in similar cases over an extended period of time (e.g., ballistics testimony, breathalyzer results, police procedures, etc.). They can also become extremely knowledgeable about the facts of each case through their involvement in pretrial proceedings. However, much of that factual knowledge (e.g., prior criminal history presented at bail hearings, motions *in limine* concerning potentially damaging evidence such as confessions) may actually bias judges' decision-making – hence, the very justification for the rules of evidence and the desirability of a trial before a jury that has not been previously exposed to that information.

Juries, on the other hand, have the advantage of pooling their collective knowledge, which maximizes their recollection and comprehension of the evidence and minimizes the possibility that biases held by any one juror will inappropriately skew the jury's interpretation of evidence. Hence, the origin of the adage that twelve heads are better than one. One set of estimates that Spencer offers assumes that judges err in their verdicts as often as juries.¹¹ Moreover, judges and juries might agree on a verdict, and both might be wrong.¹²

Based on his models, Spencer makes several findings. First, he concludes that the judge-jury agreement rate (77% in the NCSC



ON SECOND GUESSING JURY VERDICTS

study, 80% in the Kalven and Zeisel study) is only marginally better than chance.¹³ That is, if the judge and jury independently and randomly decided to convict or acquit, regardless of the evidence presented at trial, they would still agree on 62% of their verdicts. More to the point, Spencer estimates that the verdicts rendered by juries were "correct" in only 87% of the NCSC trials and only 89% of the Kalven and Zeisel trials.¹⁴

He also points out that judges and juries do not necessarily make the same types of mistakes. A key finding is that juries appear to be less likely than judges to convict an innocent defendant.¹⁵ According to Spencer, if a person is innocent, the probability that a jury will convict is 25%, while if the person is guilty, the probability that a jury will acquit is 14%. But Spencer also estimates the likely accuracy of a judge deciding the case. Assuming that the judge is at least as accurate as the jury, he estimates that if a person is innocent, the probability that a judge will convict is 37%; if the person is guilty, the probability a judge will wrongly acquit is 2%.¹⁶ Observing that many more jury verdicts are convictions than acquittals, these estimates, if accurate, would lead to the conclusion that the actual number of guilty defendants who are acquitted exceeds the number of innocent defendants who are convicted.¹⁷

Conceptual Problems with Spencer's Models of Jury Accuracy

Perhaps the most basic problem

with Spencer's model of jury accuracy is the proposition that if a judge and jury disagree on the appropriate verdict in a case, one of them is necessarily wrong. There are several inherent flaws in the assumptions underlying that proposition. First and foremost, it seems to ignore the very role of a trial in the contemporary American justice system, which is to provide a fair process in which to definitively resolve an otherwise intractable dispute. A basic premise of our justice system is that if the trial is procedurally fair, the outcome of that process is *presumed* to be correct. The presumption is intentionally designed to prevent the potential for endless re-litigation of cases.

It is not an irrebuttable presumption, however. Our justice system recognizes the possibility of trial error and includes an elaborate appeals process to correct errors when they occur. But to overcome the presumption, a litigant typically must show that the trial process was in some way procedurally flawed (e.g., by presenting inadmissible or barring admissible evidence, by prejudicial statements by the judge or counsel, etc.) or that the evidence presented at trial was legally insufficient to support the verdict. The mere fact that the trial judge would have decided the case differently, without some finding of procedural error during the trial itself, is not by itself a sufficient legal basis for concluding that either the judge or the jury was mistaken.

It is also important to recognize that jury verdicts themselves are simply the tip of the proverbial iceberg of court dispositions. Only a very small proportion of cases filed in court go to trial. And in the vast majority of instances, the reason they do so is because the facts and law are subject to different interpretations on which reasonable people can disagree. Granted, in criminal trials, the primary disagreement may be whether the prosecution can present sufficient evidence to prove its case beyond a reasonable doubt, rather than whether the defendant actually committed the crime with which he is charged. Similarly, the primary disagreement in a civil trial may be whether the plaintiff can persuade the fact-finder (judge or jury) that he was injured through the defendant's negligence or breach of contract, regardless of

whether this in fact happened. The point is that cases in which the facts and law are fairly clear – that is, there is an objective “truth” that is readily apparent to all concerned – do not go to trial; the parties resolve those cases among themselves as non-trial dispositions (dismissals, plea agreements, settlements). As a general matter, only those cases that defy resolution by any other means are ultimately tried to a jury. They are inherently difficult, complex, and subject to great differences of opinion concerning the appropriate outcome.

The NCSC trials on which Spencer bases his conclusions actually bear this out. The cases on which the judge and jury disagreed involved more criminal charges and had higher levels of evidentiary and legal complexity. The jurors themselves had more diverse opinions about how close the case really was, their initial votes during deliberations were more divided, and the deliberations were more conflicted. When the cases that ultimately go to trial provide more grounds for disagreement, we should not be surprised that the judge and jury themselves disagree in some of those cases.

Spencer’s assertion that the application of an “omniscient” or a “procedural” perspective does not affect the estimates of judge or jury accuracy exacerbates the problem. Suppose, for example, that a judge knows that the defendant has twice been convicted of armed robbery and would convict on the current charge. The prosecution presents a weak case and the jury acquits. The judge may be factually correct – that is, the defendant did commit the robbery – and jury may be procedurally correct. Now suppose that an informant testifies against the defendant, claiming that the defendant committed the robbery. The judge, familiar with informant testimony, is persuaded. The jury, finding the informant untrustworthy, acquits. The defendant actually did commit the robbery. Was the jury wrong? Or did the jury appropriately refuse to convict in the absence of strong evidence of guilt?

Finally, restricting the definition of “correct” to these two perspectives ignores the desirability that verdicts be just as well as procedurally or omnisciently correct. In some instances, juries acquit when the

***Our justice system
recognizes the
possibility of trial
error and includes
an elaborate appeals
process to correct
errors when
they occur.***

evidence strongly supports conviction (nullification) because, in the jurors’ collective judgment, the “correct” decision would lead to a miscarriage of justice.¹⁸ It is precisely the jury’s historical ability to render verdicts that reflect community beliefs about justice, mercy, and accountability that is the primary justification for trial by jury.

Inherent in the assumption that jury accuracy can be extrapolated from the judge-jury disagreement rate is the suggestion that anything less than 100% judge-jury agreement is suboptimal insofar that it implies an unacceptable number of mistaken verdicts. In making this judgment, Spencer relies on a statistical test (Cohen’s kappa) indicating that the judge-jury agreement rate in both the NCSC and the Kalven and Zeisel studies are only marginally better than chance, which he calculates at 62%.¹⁹ However, the use of that statistic is inappropriate in this instance. Spencer assumes that the pattern of higher conviction rates than acquittal is fixed. However, conviction rates vary substantially from jurisdiction to jurisdiction, ostensibly due to differences in local charging and plea policies.²⁰ A more realistic rate of “chance” judge-jury agreement would be 50 percent (as if the judge and jury independently flipped a coin to decide guilt or innocence). Compared to this

rate, an actual judge-jury agreement of 79 percent represents a substantial improvement.

Moreover, the statistical test itself was developed for use in educational testing to determine how well students would do on a test simply by randomly selecting answers to multiple choice questions, which assumes the existence of a demonstrably correct answer. Using the metrics for that statistic, Spencer suggests an inappropriate standard as the optimal level of judge-jury agreement, ignoring the historical and legal justifications for the jury trial – that is, as a bulwark against judicial tyranny, as a means to interject community values into judicial decisions, and as a means to confer public legitimacy on the judicial system and the entire governmental infrastructure.

If judges and juries always agreed on the verdict, a jury trial would be, at best, a redundant means to obtain the same decision that a judge could make with less time and expense. Although some degree of judge-jury agreement is desired – one hopes that the local bench is not totally disengaged from community values – an ideal rate has never been definitively established. As a practical matter, the optimal judge-jury agreement rate is likely to vary depending on how well the local judiciary and the local jury pool reflect the attitudinal and experiential characteristics of their respective communities across a broad range of civil and criminal cases and applicable law. Thus, judge-jury disagreement should not be interpreted as inaccuracy on either the judge’s or the jury’s part, but rather as a reflection of a healthy democratic institution working to keep the outcomes of the judicial process in sync with community mores.

This historical role of juries in the American justice raises another conceptual shortcoming in Spencer’s focus on jury accuracy – namely, that jury performance cannot and should not be considered in a vacuum. Rather, jury performance must be considered in light of the accuracy rates for available alternatives to jury trials. We abandoned the practice of trial by ordeal centuries ago because it proved to be a highly unreliable method of determining guilt or innocence. In contemporary America, many litigants opt out of traditional trial processes in favor of mediation,

arbitration, diversionary criminal programs, and other non-traditional processes that have been developed to obtain more favorable outcomes to litigants and the larger society. But for those cases that litigants fail to resolve among themselves, whether criminal or civil in nature, our justice system has essentially one alternative: a bench trial. Not only does Spencer's conceptual focus on the accuracy of jury verdicts overlook how jury performance compares to that of trial judges, his ability to quantify their respective performance is severely limited by several methodological problems in the study, to which I now turn.

Methodological Problems

Spencer acknowledges several potential methodological problems in his models – namely, survey error, specification error, and identification error – and attempts to take the impact for the latter two of those into account in his findings. But he glosses over the considerable likelihood of survey error in his calculations of judge-jury agreement, especially in the NCSC study. That study was intended as an examination of jury deadlock in felony trials in response to calls for non-unanimous verdicts. The survey questions posed to judges and jurors were not drafted with the intent to investigate the extent to which judges and juries agreed with one another on the verdict.

Information about the jury's verdict in the hung jury study was quite detailed. The NCSC obtained the precise disposition (*e.g.*, conviction, acquittal, deadlock, charge not considered by jury, unknown) for each of the charges and lesser-included charges tried to the jury. The judges, in contrast, were simply asked what their judgment would have been (conviction or acquittal) if the case had been tried as a bench trial. They did not have the option to indicate a split (conviction/acquittal) or undecided verdict in cases involving multiple or lesser-included charges in the indictment (approximately 70% of the original dataset) or multiple defendants (8% of the original dataset). Thus, the comparison between the jury's verdict and the judge's verdict preference is not neatly symmetrical. Moreover, the judge was asked to indicate his/

Even without egregious errors in the investigation and prosecution of both criminal and civil cases, the trial process itself is an inherently imperfect vehicle for fostering informed decisions.

her verdict preference before the jury actually returned its verdict (to prevent biasing the judge's response to conform to the actual verdict). Judges were not subsequently asked whether they thought the jury's verdict was factually or legally incorrect, a response that is qualitatively different from one that merely acknowledges a difference of opinion as to the inferences that can be reasonably drawn from the trial evidence.

Spencer also omitted 68 of the NCSC trials in which the jury deadlocked on one or more charges (19% of the original dataset) in his analyses. As we found in the original study of hung juries, cases in which the jury deadlocked involved significantly more complex and ambiguous evidence, greater conflict during deliberations, and increased perceptions by jurors that the law they were asked to apply was somehow unfair in light of the overall circumstances of the case. Arguably, for those cases in which the jury deadlocked on one or more charges, the judge would have agreed with at least some of the jurors. Omitting those cases from his analyses further distorts the calculations of judge-jury agreement.

To examine the extent to which evidentiary ambiguity contributes to jury deadlock, the original NCSC surveys asked the judges and jurors to indicate their respective assessments about the strength of the evidence on a scale of 1 (evidence strongly favored the prosecution) to 7 (evidence strongly favored the defense). The jurors' responses for each case were averaged into a single composite value and compared to those reported by the respective judges. Spencer used these variables to further assess judge-jury agreement, but he collapsed the 7-point scale into a 3-point scale indicating strong evidence for conviction (1 or 2 on the 7-point scale), medium evidence for conviction (3 to 5 on the 7-point scale), and weak evidence for conviction (6 or 7 on the 7-point scale).

In doing so, however, Spencer creates the perception of greater disagreement than actually occurred. For example, it is not clear that a jury's rating of 3 on the 7-point scale differs in a meaningful way from a judge's rating of a 2, which occurred in 68 of the trials (35.3%) in Spencer's analysis. Only 22% of juries and judges agreed exactly with each other's evidentiary ratings, but two-thirds of those ratings fall within one step of their counterparts'. But by collapsing the 7-point scale into a 3-point scale, more than one-third of the judges' ratings fall into a different category than the juries' ratings, although the actual difference is quite small. Sixty percent of the judicial ratings, and 82% of the jury ratings, fall on the edge of the thresholds (that is, 2 or 3, 5 or 6) for the collapsed categories.

Finally, the standard errors in Spencer's models for the marginal probability of an incorrect verdict by the jury or judge are extremely large, in some instances, larger than the actual probabilities themselves. These raise serious doubts about the accuracy of the estimates on which Spencer makes his ultimate conclusions.

Implications and Conclusions

Despite the many conceptual and methodological problems inherent in Spencer's study of jury accuracy, we are still confronted with the reality that the jury's verdict in a small number of cases has been shown to be demonstrably wrong. Investigative

efforts by the Innocence Project have resulted in the exoneration of 206 individuals who were convicted of crimes that they did not commit.²¹ These, plus the small proportion of civil jury verdicts that are reversed in JNOV or appellate decisions due to insufficient evidence to support the verdict, are concrete examples of "incorrect verdicts" that prove that juries sometimes do err, even if we cannot determine with any certainty how often they do so.

Even while conceding this point, a closer look at the Innocence Project and other exoneration cases reveals that incorrect verdicts are not so much the result of mistaken decision-making by the jury, but rather of a flawed criminal justice process of which the jury trial was just one segment. Studies of wrongful conviction repeatedly find that the major causes include incorrect eyewitness identification, faulty forensic evidence, unreliable informant testimony, and false confessions.²² In other words, these cases are caused by errors in police and prosecutor procedures that identify, focus on, and bring the wrong defendant to trial, while failing to provide jurors with information that might cause them to question the reliability or

credibility of the evidence.

Even without egregious errors in the investigation and prosecution of both criminal and civil cases, the trial process itself is an inherently imperfect vehicle for fostering informed decisions. Juries in both criminal and civil trials are frequently presented with incomplete or imperfect evidence, and the quality of trial presentations vary substantially depending on the skills of the witnesses, lawyers, and the judge. It is simply inaccurate to place the responsibility for these mistaken verdicts on the juries who decided those cases, or even on the judges who presided in these trials. Spencer's attempt to quantify jury accuracy diverts our attention from problems in our justice system that contribute more substantially and directly to unjust outcomes than mistaken judgments by jurors. Rather, it seems that the jury trial may be the best, and in many cases the last, opportunity for the justice system to produce a just and accurate outcome from an imperfect process.

Rather than bemoaning the fact that the jury is sometimes unable to correct the mistakes that preceded trial, we should concentrate on providing jurors with all of the information and tools they might need to improve

the likelihood that they will be able to recognize and appropriately weigh incomplete, imperfect, or ambiguous evidence. For example, judges should permit prosecutors and defense counsel to present evidence about the police procedures employed in obtaining eyewitness identifications and the error rates associated with those methods.²³ From the moment jurors receive their summonses and report to the courthouse, their treatment by judges, court staff, and lawyers, should impress upon them the importance and seriousness of their responsibilities and inspire them to put every effort into setting aside preconceptions and potentially biasing attitudes and into judging the evidence fairly and objectively.²⁴ Finally, jurors should be permitted to take notes during trial, should be permitted to submit questions to witnesses, should be instructed about the applicable law before the evidentiary portion of the trial, and should be provided with written copies of the final jury instructions for use during deliberations. These and other techniques improve juror comprehension and performance, reduce the possibility of juror error, and improve the confidence of judges, lawyers, litigants and the public in the legitimacy of the jury's verdict.²⁵

* Director, Center for Jury Studies, National Center for State Courts.

¹ *New Study Shows Juries Get It Wrong* (Northwestern University Institute for Policy Research, press release issued June 19, 2007).

² Bruce D. Spencer, *Estimating the Accuracy of Jury Verdicts*, J. EMPIRICAL LEG. ST. (forthcoming).

³ *Id.* at 10. Because the cases were not randomly selected from among all of cases tried in those courts, Spencer cautions that the results of his study cannot be generalized to all jury trials in all locations.

⁴ A larger proportion of jury verdicts in the sample resulted in convictions. So, even though the probability of wrongful conviction was greater than that of a wrongful acquittal, the actual number of guilty defendants that are acquitted exceeds the number of innocent defendants that are convicted. *Id.* at 18-19.

⁵ *Id.* at 2.

⁶ See Paula L. Hannaford-Agor, Valerie P. Hans, Nicole L. Mott & G. Thomas Munsterman, *Are Hung Juries A Problem? Final Report to the National Institute of Justice* (Sept. 30, 2002).

⁷ HARRY KALVEN & HANS ZEISEL, *THE AMERICAN JURY* (1966).

⁸ Spencer, *supra* note 2, at 29-30.

⁹ at 6-8.

¹⁰ *Id.* at 2.

¹¹ *Id.* at 7-8.

¹² *Id.*

¹³ *Id.* at 5.

¹⁴ *Id.* at 10.

¹⁵ *Id.*, Table 4, at 20.

¹⁶ *Id.*

¹⁷ *Id.* at 18-19.

¹⁸ Paula L. Hannaford-Agor & Valerie P. Hans, *Nullification at Work? A Glimpse from the National Center for State Courts Study of Hung Juries*, 78 CHI.-KENT L. REV. 1249 (2003).

¹⁹ The formula is the judge acquittal rate times the jury acquittal rate plus the judge conviction rate times the jury conviction rate (32.1% * 16.5% + (67.9% times 83.5%). Spencer, *supra* note 2, at n.6.

²⁰ See, e.g., Hannaford-Agor et al., *supra* note 6, at 19-20.

²¹ See <http://www.innocenceproject.org/>.

²² Brandon L. Garrett, *Judging Innocence*, COLUMBIA L. REV. (forthcoming 2008); Samuel R. Gross et al., *Exonerations in the United States 1989 Through 2003*, 95 J. CRIM. L. & CRIMINOLOGY 523 (2005).

²³ For a summary of the literature on this topic, see the U.S. DEPARTMENT OF JUSTICE, OFFICE OF JUSTICE PROGRAMS, EYEWITNESS EVIDENCE: A GUIDE FOR LAW ENFORCEMENT, RESEARCH REPORT (1999).

²⁴ Paula L. Hannaford-Agor & G. Thomas Munsterman, *Ethical Reciprocity: The Obligations of Citizens and Courts to Promote Participation in Jury Service*, in JOHN KLEINIG & JAMES P. LEVINE (eds.), JURY ETHICS: JUROR CONDUCT AND JURY DELIBERATIONS 21 (2006).

²⁵ G. THOMAS MUNSTERMAN, PAULA L. HANNAFORD-AGOR, & G. MARC WHITEHEAD, *JURY TRIAL INNOVATIONS* (2d ed. 2006); AMERICAN BAR ASSOCIATION, *PRINCIPLES FOR JURIES AND JURY TRIALS* (2005).