Cornell Law Library
Scholarship@Cornell Law: A Digital Repository

Cornell Law Faculty Publications | Faculty Scholarship

3-1-2005

Judge-Jury Agreement in Criminal Cases: A Partial Replication of Kalven and Zeisel's The American Jury

Theodore Eisenberg
Cornell Law School, ted-eisenberg@lawschool.cornell.edu

Paula L. Hannaford-Agor
National Center for State Courts, phannaford@ncsc.dni.us

Valerie P. Hans
Cornell Law School, valerie.hans@cornell.edu

Nicole L. Waters
National Center for State Courts, nwaters@ncsc.org

G. Thomas Munsterman
National Center for State Courts, tmunsterman@ncsc.dni.us

See next page for additional authors

Follow this and additional works at: http://scholarship.law.cornell.edu/facpub

Part of the Applied Statistics Commons, Criminal Law Commons, Criminal Procedure Commons, and the Litigation Commons

Recommended Citation

Eisenberg, Theodore; Hannaford-Agor, Paula L.; Hans, Valerie P.; Waters, Nicole L.; Munsterman, G. Thomas; Schwab, Stewart J.; and Wells, Martin T., "Judge-Jury Agreement in Criminal Cases: A Partial Replication of Kalven and Zeisel's The American Jury" (2005). *Cornell Law Faculty Publications.* Paper 343.
http://scholarship.law.cornell.edu/facpub/343

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Faculty Publications by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

Authors
Theodore Eisenberg, Paula L. Hannaford-Agor, Valerie P. Hans, Nicole L. Waters, G. Thomas Munsterman, Stewart J. Schwab, and Martin T. Wells

This article is available at Scholarship@Cornell Law: A Digital Repository: http://scholarship.law.cornell.edu/facpub/343

Journal of Empirical Legal Studies
Volume 2, Issue 1, 171–206, March 2005

Judge-Jury Agreement in Criminal Cases: A Partial Replication of Kalven and Zeisel's *The American Jury*

*Theodore Eisenberg, Paula L. Hannaford-Agor, Valerie P. Hans, Nicole L. Waters, G. Thomas Munsterman, Stewart J. Schwab, and Martin T. Wells**

This study uses a new criminal case data set to partially replicate Kalven and Zeisel's classic study of judge-jury agreement. The data show essentially the same rate of judge-jury agreement as did Kalven and Zeisel for cases tried almost 50 years ago. This study also explores judge-jury agreement as a function of evidentiary strength (as reported by both judges and juries), evidentiary complexity (as reported by both judges and juries), legal complexity (as reported by judges), and locale. Regardless of which adjudicator's view of evidentiary strength is used, judges tend to convict more than juries in cases of "middle" evidentiary strength. Judges tend to acquit more than juries in cases in which judges regard the evidence favoring the prosecution as weak. Judges tend to convict more than juries in cases in which judges regard the evidence favoring the prosecution as strong. Rates of adjudicator agreement are thus partly a function of which adjudicator's view of evidentiary strength is used, a result not available to Kalven and Zeisel, who were limited to judges' views of the evidence. We find little evidence that evidentiary complexity or legal complexity help explain rates of

---

*Address correspondence to Theodore Eisenberg, Cornell Law School, Myron Taylor Hall, Ithaca, NY 14853; email: theodore-eisenberg@postoffice.law.cornell.edu. Theodore Eisenberg is the Henry Allen Mark Professor of Law, Cornell Law School; Paula L. Hannaford-Agor is Principal Court Research Consultant, National Center for State Courts (NCSC); Valerie P. Hans is Professor of Criminal Justice and Psychology at the University of Delaware; Nicole L. Waters is Court Research Associate, NCSC; G. Thomas Munsterman is Principal Court Management Consultant, NCSC; Stewart J. Schwab is Allan R. Tessler Dean, Cornell Law School; Martin T. Wells is Professor of Statistics, Department of Social Statistics, Cornell University.

Drafts of this article were presented at the 2003 Law and Society Association Meeting, Pittsburgh, at the 2003 American Law and Economics Association Meeting, Toronto, and at the 2004 Meeting of American Psychology-Law Society, Scottsdale. The gathering of the data used in this article was supported by Grant No. 98-IJ-CX-0048, National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice or the National Center for State Courts.

judge-jury disagreement. Rather, the data support the view that judges have a lower conviction threshold than juries. Local variation exists among the sites studied. The influences of juror race, sex, and education are also considered.

In a study of 1950s' criminal trials, Harry Kalven, Jr. and Hans Zeisel, relying on questionnaire responses from judges, found substantial agreement between how judges and juries would rule in the same case.[1] They also found that disagreement was asymmetric. Juries tended to acquit when judges would convict much more than juries tended to convict when judges would acquit.[2] Kalven and Zeisel attributed most instances of judge-jury disagreement to different interpretations of the reasonable-doubt standard, different evaluation of the evidence, juror sentiments about the defendant, and juror disagreement with the law.[3] They believed that such factors influenced the jury when the judge regarded the case as close and capable of supporting either conviction or acquittal.[4] They also concluded that judge-jury disagreements were not due to jury misunderstanding of the evidence.[5] Kalven and Zeisel's work is acknowledged to be the first large-scale systematic study of the jury and remains in many scholars' minds the most significant.[6] "Now

---

[1]Harry Kalven, Jr. & Hans Zeisel, The American Jury 58 (1966).

[2]Id. See Valerie P. Hans & Neil Vidmar, The Twenty-Fifth Anniversary of *The American Jury*, 16 Law & Soc. Inquiry 323, 327 (1991).

[3]Kalven & Zeisel, supra note 1, at 106–11. Minor factors explaining judge-jury disagreement were facts known only to the judge and disparity in the quality of counsel. Id. at 111. See Hans & Vidmar, supra note 2, at 328 (summarizing Kalven and Zeisel's explanations).

[4]Kalven & Zeisel, supra note 1, at 106, 185; Hans & Vidmar, supra note 2, at 328.

[5]Kalven & Zeisel, supra note 1, at 149; Hans & Vidmar, supra note 2, at 328.

[6]Hans & Vidmar, supra note 2; Valerie P. Hans, Paula L. Hannaford-Agor, Nicole L. Mott & G. Thomas Munsterman, The Hung Jury: *The American Jury*'s Insights and Contemporary Understanding, 39 Crim. L. Bull. 33 (2003); Wayne A. Logan, Symposium: Reflections on Kalven and Zeisel's *The American Jury*, 39 Crim. L. Bull. 4, 5 (2003) (referring to "the monumental significance of Kalven and Zeisel's work, which for the first time in history subjected the American criminal jury to comprehensive description and analysis").

over thirty-five years old, *The American Jury* remains the benchmark in U.S. jury research."[7]

This article uses a new criminal case database assembled by the National Center for State Courts (NCSC) to partially replicate the Kalven and Zeisel study. Based on more than 300 trials in four locales, we find the rate of judge-jury agreement, over 70 percent, to be substantially the same as that found by Kalven and Zeisel. We also find a similar asymmetry in the pattern of disagreement; juries are much more likely to acquit when judges would convict than they are to convict when judges would acquit.

Following Kalven and Zeisel's effort to explain the pattern of disagreement between judges and juries, we study the pattern as a function of evidentiary strength, evidentiary complexity, and legal complexity. The new data set provides information about evidentiary strength that was not available to Kalven and Zeisel. The data include the views of four different categories of participants—judges, jurors, prosecutors, and defense attorneys—about the closeness of the case. Using these different perspectives about the cases, we find evidence that judges tend to convict more than juries. This finding addresses one of the themes of criticism of Kalven and Zeisel—that their analysis depended too much on what judges said about what jurors thought.[8] This study has the reactions of the other actors in the system as well as judges' reactions.

We also find that the judicial tendency to convict, relative to the jury, is not universal. When judges regard evidence of conviction as weak, they tend to acquit more than juries. The tendency to convict more than juries does not emerge in all four locales studied. We find little evidence that evidentiary complexity or legal complexity help explain rates of judge-jury disagreement. The judges' lower conviction threshold seems to be driving most of the difference.

Section I of this article describes the NCSC data used in this study. Section II reports on rates of judge-jury disagreement. Section III explores the sources of judge-jury disagreement by focusing on one explanatory factor at a time. Section IV combines the factors discussed in Section III in regression models.

---

[7]Logan, supra note 6, at 5.

[8]Abraham S. Goldstein, Book Review, 1 Law & Soc'y Rev. 148, 151 (1967); Hans & Vidmar, supra note 2, at 330, 334. See Kalven & Zeisel, supra note 1, at 50.

A national replication of Kalven and Zeisel's work was called for over a decade ago.[9] It has been difficult to achieve because it requires substantial funding and sustained collaboration of busy, real-world actors and interested researchers. We acknowledge the limitations inherent in using just four sites in the current analyses. However, replication of Kalven and Zeisel's important work has to start somewhere. We return to this issue in the conclusion.

I. The Data

The National Center for State Courts gathered the data used in this study as part of a project, funded by the National Institute of Justice (NIJ), to study hung juries. The NCSC's report, *Are Hung Juries a Problem?*, thoroughly describes the data,[10] so the description here, derived from the NCSC report, is more abbreviated.

Four sites participated in the data collection: the Central Division, Criminal, of the Los Angeles County Superior Court, California; the Maricopa County Superior Court (Phoenix), Arizona; the Bronx County Supreme Court, New York; and the Superior Court of the District of Columbia. Several criteria shaped site selection. First, each site needed a sufficiently high volume of felony jury trials to permit data collection within a reasonable time period. Second, court personnel had to be willing to cooperate in data collection, including agreeing to adhere to privacy and confidentiality protocols. Los Angeles and Washington, DC were included because of reported concerns about hung-jury rates. Maricopa was chosen to study the effects of an innovative procedure allowing judges to admit further evidence and arguments in cases with deadlocked juries. The New York State Office of Court Administration provided suggestions about high-volume courts in New York City, and helped secure the Bronx County Supreme Court's cooperation.

---

[9]Hans & Vidmar, supra note 2, at 347–48.

[10]Paula L. Hannaford-Agor, Valerie P. Hans, Nicole L. Mott & G. Thomas Munsterman, Are Hung Juries a Problem? (National Center for State Courts Sept. 30, 2002) [hereafter "Hung Juries"], available at <http://www.ncsconline.org/WC/Publications/BesJuriesHungJunesPub.pdf>. See also Hans et al., supra note 6.

After a pretest in Los Angeles,[11] data were collected at the four sites. Data were collected in Los Angeles from June 2000 through October 2000.[12] Maricopa County began data collection in November 2000 and ended in October 2001.[13] Data from the Bronx were collected from February through August 2001, and data for Washington, DC were collected from April through August 2001.

Court personnel at the sites distributed and collected questionnaire packets covering noncapital felony cases in all site courtrooms. The sample excludes misdemeanor cases because hung juries in felony trials are typically of greater concern to policymakers. The sample excludes capital cases because of the sanction's severity and because of the risk that confidential juror questionnaire data might be used in litigation. Further information about the data is available in NCSC's report.[14]

A. The Questionnaires

Each packet contained instructions and questionnaires for the judges, attorneys, and jurors. Each packet also had a case data form requesting information about case characteristics and outcomes. Many of the questions asked trial participants to give ratings on a seven-point Likert scale. The content of each questionnaire most relevant to this study is listed below, and is described more fully in the NCSC report.[15]

- *Case Data Form*—type of charge, sentence range, jury's decision, and demographic information about the defendant(s) and the victim(s), voir dire, trial evidence and procedures, and jury deliberations.

---

[11]Hung Juries, supra note 10, at 29.

[12]This period overlapped with a significant local investigation of the Los Angeles police (the Ramparts investigation), producing some concern about the typicality of the conviction/acquittal ratio.

[13]A brief hiatus during this period was the result of some confusion on which cases were to be included in the study. For a short time, some judges believed that data were to be collected only if the jury hung. Thus, the number of hung juries in Maricopa County may be higher than is typically the case. The hung-jury rates in Maricopa were 3.3 percent (hung on all counts), 5.1 percent (hung on Count 1), and 7.7 percent (hung on any count).

[14]Hung Juries, supra note 10.

[15]Id. at Appendix A.

- *Judge Questionnaire, Part I: Before Jury Verdict*—verdict judge would have reached in bench trial, evaluation of the evidence, case complexity, attorney skill, likelihood that the jury would hang.
- *Judge Questionnaire, Part II: After Jury Verdict*—reaction to the verdict, and experience on the bench.
- *Attorney Questionnaire, Part I*—assessment of voir dire, case complexity, attorney skill, evaluation of the evidence.
- *Attorney Questionnaire, Part II*—reaction to the verdict, and experience in legal practice.
- *Juror Questionnaire*—case complexity, attorney skill, evaluation of the evidence, formation of opinion, the dynamics of the deliberations including the first and final votes, juror participation, conflict, reaction to verdict, opinion about applicable law, assessment of criminal justice in the community, and demographic information.

B. Distribution of Study Packages

Researchers briefed judges and key court personnel about the project and instructed about how the packet distribution was to occur.[16] Packets were sent from the jury assembly room to the courtrooms with the panel for voir dire. Once the jury was selected, court personnel distributed the packets to the judge and/or court clerk. If the case proceeded through to jury deliberations, and did not end by a plea agreement, dismissal, or mistrial for some reason other than the jury's inability to arrive at a unanimous verdict, the judge was asked to complete the judge survey. In addition, either the clerk or the judge was to complete a questionnaire on the general case information on a case data form.

Once the jury retired to deliberate, court personnel distributed the judge and attorney questionnaires. The judges and attorneys were asked to complete the questionnaires in two stages, answering some questions prior to the jury decision (Part I) and the remaining questions after the jury rendered its verdict or the case declared a mistrial (Part II). The court personnel distributed the final set of questionnaires to the jurors after the verdict was announced or a mistrial declared. To protect confidentiality, respondents were provided with a blank envelope in which to place the completed

---

[16]See Hung Juries, supra note 10, at 33 (fig. 3.1), for the time sequence of the packet distribution.

questionnaire. Court staff collected the completed questionnaires and gave these to the designated court liaison for each site, who forwarded the cases to the NCSC for data entry and analysis.

C. Response Rates

The NCSC report summarizes response rates.[17] Briefly, case data forms were returned in 358 of the 401 cases, an 89 percent response rate. Judges completed 366 (91 percent response rate) questionnaires. Although confidentiality precludes us from linking the data in a particular case to an individual judge, we are confident that a substantial number of judges are represented in the sample. For example, in Maricopa County, 29 judges sat in the criminal division in fiscal year 2000–2001.[18]

There were 576 total attorney questionnaires (either defense or prosecution) completed in 351 cases. At least one attorney responded in 88 percent of the cases and the prosecutor and defense counsel both responded in 64 percent of the cases. At least one defense attorney completed a form in 278 cases (69 percent response rate) and at least one prosecuting attorney in 287 cases (72 percent response rate). Overall, 3,626 jurors returned their questionnaires. The response rate for jurors across all sites, with consideration for jury size, was 80 percent. For the 12-person juries in Los Angeles, Maricopa, the Bronx, and DC, the average response rate was 11, 10, 8, and 10 jurors, respectively. For the eight-person juries in Maricopa,[19] an average of seven jurors responded. The case data form surveys asked factual information about the criminal charges filed and the jury's decision,[20]

---

[17]Id. at 32.

[18]Email from Judge G. T. Anagnost to Theodore Eisenberg, Jan. 26, 2004. See also note 61 infra.

[19]In California, New York, and DC, felony cases are tried to a 12-person jury. Arizona law provides for eight-person juries in felony trials unless the penalty for the defendant includes death or a potential sentence of 30 years or more, in which case the number of jurors is 12. In Maricopa County, there were 30 cases with 12-member juries. Sixty-nine juries had eight members. In addition, there were six cases with so little information that jury size could not be determined.

[20]If this key information was missing from the questionnaires, NCSC made follow-up inquiries with the courts. Thirty-one cases without a case data form were salvaged through direct communication with the courts to obtain the key information about the case. The courts were unable to recover this missing information in 12 cases, which were not included in the final

and led to 382 usable cases. The number of questionnaires included in the final usable database varied slightly for each site and are summarized in the NCSC report.[21]

D. Comments on the Data

The NCSC data have important limitations. A comparative weakness is a smaller, less geographically diverse sample than Kalven and Zeisel had, though even Kalven and Zeisel's sample was criticized as not being representative.[22] The NCSC data design actively sought locales with enough hung juries to support a study.[23] Like Kalven and Zeisel, one cannot be sure that the judges in the NCSC sample would have decided to acquit or convict had the judge actually been adjudicating the case.[24] The jury, in contrast, always made the decision with real-world impact.

The NCSC data were gathered in a legal environment that differs from Kalven and Zeisel's, thereby perhaps limiting the range of expected judge-jury agreement. Juries have become more representative of the community than they were in the 1950s. Judges have changed as well, but the net change in jury composition likely is greater. "Although the judiciary has also become more democratized, it still represents a much narrower segment of society than the jury. If life experiences of decision makers influence their judgments, juries and judges might disagree *more* now than they did in the 1950s when Kalven and Zeisel collected their data."[25] In addition, the gap between defense and prosecutorial resources may have narrowed since the 1950s. At

---

data analysis. Seven additional cases had so little information (three or fewer questionnaires received) that they were also eliminated from the analysis. For example, a few cases relied on only one questionnaire from one attorney and were therefore not included.

[21]Hung Juries, supra note 10, at 33 (tbl. 3.1).

[22]Kalven & Zeisel, supra note 1, at 36; Hans & Vidmar, supra note 2, at 331 (describing review of Kalven and Zeisel's book appearing in the *American Sociological Review*).

[23]Hung Juries, supra note 10, at 29.

[24]Kalven & Zeisel, supra note 1, at 53–54; Goldstein, supra note 8, at 151; Hans & Vidmar, supra note 2, at 331.

[25]Hans & Vidmar, supra note 2, at 348 (footnote omitted).

a minimum, *Gideon v. Wainwright*'s[26] requirement of counsel in serious criminal cases (which brought a few states into line with others' practices), and *Miranda v. Arizona*'s[27] requirement that defendants be told of this right, might have effects throughout the criminal process, including trial outcomes.

The NCSC data also offer some advantages in explaining results that were not available to Kalven and Zeisel. The NCSC data design assured that views of evidence and complexity other than those of the judge were available, thus reducing an important concern about Kalven and Zeisel's analysis—that the juries' attitudes were seen only through the eyes of the judge. Kalven and Zeisel acknowledged "an inability to know for certain that judges were providing accurate insights into jury behavior rather than stereotypical judicial interpretations."[28] In addition, the NCSC data are available at a time when more sophisticated statistical techniques are available. Statistical concerns about Kalven and Zeisel's analysis include their failure to supply significance levels[29] and their tendency to reach conclusions based on small subsamples of their data.[30] Statistical techniques now more widely used, especially for the categorical data comprising the sample, facilitate studying multiple factors simultaneously, thereby enriching the analysis.

II. Rates of Agreement

Both replication and enhancement of Kalven and Zeisel's study are crucial to determine (1) whether Kalven and Zeisel's use of the judge's view of jury behavior unduly shaped their explanations of disagreement, (2) whether

---

[26]372 U.S. 335 (1963) (indigent defendant has right to counsel).

[27]384 U.S. 436, 478–79 (1966) (statements obtained from criminal detainees, without warning of constitutional rights, held inadmissible).

[28]Hans & Vidmar, supra note 2, at 334.

[29]A. Paul Hare, Book Review, 32 Am. Soc. Rev. 666, 667 (1967); Herbert Jacob, Book Review, 11 Midw. J. Pol. 282, 283 (1967).

[30]John Kaplan, Book Review, 115 U. Pa. L. Rev. 475, 477–78 (1967). Kalven and Zeisel acknowledged that analyzing subgroups of their data "exhausted our sample." Kalven & Zeisel, supra note 1, at 91.

changes over time in juror pools, the flow of cases to the jury, or the criminal justice system alter their basic findings, and (3) to help resolve inconsistencies between Kalven and Zeisel's findings and some experimental results.[31]

A. Compared to Kalven and Zeisel

To facilitate comparison of judge-jury agreement rates in the NCSC data with the rates reported by Kalven and Zeisel, we initially follow the Kalven and Zeisel decision to allocate hung juries evenly across convictions and acquittals.[32] Panel A of Table 1 reports the agreement rates for the NCSC data.[33] This table is directly comparable to Kalven and Zeisel's famous Table 12, which is reproduced as Panel B of Table 1.

Table 1's most striking feature is the similarity of Panels A and B. The total rate of agreement in Panel A, computed by adding the percentage of cases in which judges and juries agreed to acquit or convict is 75 percent. The total rate of agreement in Panel B is 78 percent. Even with the large number of observations, we cannot reject, at conventional levels of statistical significance ($p = 0.160$), the hypothesis that the agreement rates are the same.

---

[31]Experimental research, for example, suggests that compared to judges, jurors may have lower rather than higher standards of reasonable doubt. Dorothy K. Kagehiro & W. Clark Stanton, Legal vs. Quantified Definitions of Standards of Proof, 9 Law & Hum. Behav. 159 (1985) (college students' mock verdicts not affected by different legal standards). Experimental research also suggests that jurors have difficulty in distinguishing standards of proof. Simon and Mahon asked 106 judges and 25 jurors to convert the "reasonable doubt" and "preponderance of the evidence standards" into numerical terms. Rita James Simon & Linda Mahan, Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom, 5 Law & Soc'y Rev. 319 (1971). Judges distinguished between the two standards, requiring an 89 percent probability of an event for the "reasonable doubt" standard and 61 percent for the civil standard. Jurors did not distinguish between the two standards, and reported probabilities of 79 percent and 77 percent for the two standards, respectively. Kassin and Wrightsman point out that juries are unlikely to be as oblivious to burdens of proof as experimental research suggests. The pattern of jury behavior in experimental work suggests that, relative to the judge, jurors should favor the prosecution in criminal trials. Saul M. Kassin & Lawrence S. Wrightsman, The American Jury on Trial: Psychological Perspectives 156 (1988). Yet Kalven and Zeisel's data suggest the opposite, finding juries more likely to exhibit leniency in criminal cases. Id.

[32]Kalven & Zeisel, supra note 1, at 58.

[33]In 79 cases judges did not provide a hypothetical verdict, and in an additional 28 cases there were mixed jury verdicts that could not be classified as predominantly convictions or acquittals. Hans et al., supra note 6, at 56 n.129.

Table 1: Judge-Jury Agreement on Conviction

	Jury Acquits	*Jury Convicts*
A. NCSC Data		
Judge acquits	13%	6%
Judge convicts	19%	62%
B. Kalven & Zeisel Data		
Judge acquits	14%	3%
Judge convicts	19%	64%

NOTE: Panel A is based on the NCSC data covering 290 nonhung-jury trials with both judge and jury data, and 28 hung-jury trials, at four sites in 2000–2001. Panel B is based on the Kalven and Zeisel data covering 3,576 trials before 555 judges in 47 states and the District of Columbia in 1954, 1955, and 1958.

Panel A's asymmetry of disagreement also matches Kalven and Zeisel's data. Jurors in both data sets are much more likely to disagree with a judge's decision to convict than with a judge's decision to acquit. A McNemar test of the significance of the asymmetry of the disagreement pattern in Panel A yields a p value of less than 0.0001 and the p value for Panel B is vanishingly small. Thus, looking at the cases of disagreement between judge and jury, it is extremely unlikely that the higher conviction rate by judges in cases of disagreement would emerge by chance. We do note, however, a greater tendency of jurors in the NCSC sample to convict when the judge would acquit than was present in the Kalven and Zeisel sample. Limiting the sample to cases of disagreement, this increase, from 14 percent of disagreement cases in Kalven and Zeisel's data (3 percent out of the 22 percent of cases in which disagreement occurred), to 24 percent of disagreement cases in the NCSC data (6 percent out of the 25 percent of cases in which disagreement occurred), is statistically significant ($p = 0.018$).[34]

Kalven and Zeisel's Table 12 was a simplification of their Table 11, in which they separately reported rates of hung juries, rather than allocate hung-jury cases equally between convictions and acquittals as in their Table 12.[35] Table 2 reproduces Kalven and Zeisel's Table 11 and shows the same

[34]If one excludes hung-jury cases, based on the figures reported in Table 2, the difference is more significant ($p = 0.007$).

[35]Kalven & Zeisel, supra note 1, at 56. Other protocols for reporting disagreement could be followed. Kalven and Zeisel later report agreement rates that account for the fact that some cases involve multiple counts and that judges and juries could agree to convict but disagree on the

Table 2: Judge-Jury Agreement, Including Hung-Jury Cases

	Jury Acquits	*Jury Convicts*	*Hung Jury*
A. NCSC Data			
Judge acquits	11.6%	5.0%	1.9%
Judge convicts	16.0%	58.5%	6.9%
B. Kalven & Zeisel Data			
Judge acquits	13.4%	2.2%	1.1%
Judge convicts	16.9%	62.0%	4.4%

NOTE: Panel A is based on the NCSC data covering 290 nonhung-jury trials with both judge and jury data, and 28 hung-jury trials, at four sites in 2000–2001. Panel B is based on the Kalven and Zeisel data covering 3,576 trials before 555 judges in 47 states and the District of Columbia in 1954, 1955, and 1958.

information for the NCSC data. Excluding hung-jury cases, Kalven and Zeisel show judge-jury agreement in 75.4 percent of cases and the NCSC data show judge-jury agreement in 70.5 percent of cases. The difference, while not statistically significant, is somewhat larger than that suggested in following their decision to evenly allocate hung-jury cases.

Table 2 also shows that the hung-jury rate is somewhat higher in the NCSC data (8.8 percent) than in the Kalven and Zeisel data (5.5 percent). This is not surprising because some jurisdictions were selected for study because of a belief that they might have high hung-jury rates.[36] Even so, the 3.3 percent difference in hung-jury rates is probably a bit high. Kalven and Zeisel report that their data may understate the true rate of hung juries because of their judges' decisions about what constituted a reportable trial.[37] Some trials ending in hung juries might have gone unreported. Other rea-

count. Judge and jury agreed to convict on the same count in 56.8 percent of cases. Id. at 60 (tbl. 13). Adding this figure to the 13 percent of cases in which they agreed to acquit yields an agreement rate of 69.8 percent. See Hung Juries, supra note 10, at 55–56, for a measure of agreement that accounts for multiple charge cases.

[36]Hung Juries, supra note 10, at 29 ("Two sites were included because of reported concerns about hung jury rates (Los Angeles and Washington, DC).").

[37]Kalven & Zeisel, supra note 1, at 57 ("Since our instructions to the judges as to what constitutes a reportable jury trial were perhaps imprecise on the point, it is possible that some felt no need to report on what is technically a mistrial.").

sonably systematic evidence supports the notion that Kalven and Zeisel's data understate modern hung-jury rates.[38]

B. Variation Across Sites

1. Agreement Rates

The four NCSC sites vary in their rates of agreement, as described in Table 3.[39] The pattern of agreement varies both in magnitude and overall direction. The rate of judge-jury agreement varies from 64 percent in DC to 89 percent in Maricopa. Do these results meaningfully differ from Kalven and Zeisel? We cannot know for certain because Kalven and Zeisel do not report site-specific rates of judge-jury agreement. In an appendix, however, they describe a few other studies reporting site-specific rates of agreement, ranging from 72 percent in civil cases to 89 percent in criminal cases.[40] We suspect that Kalven and Zeisel's aggregate rates of agreement of over 70 percent are likely the result of locales in which agreement rates were less than 70 percent and in which agreement rates exceeded 80 percent, as well as locales in which agreement rates were in the 70 percent range. If that is so, then the NCSC data may not materially differ. More definitive word on local variation must await samples that encompass additional jurisdictions. For now, we merely note that Kalven and Zeisel's aggregated reporting of results may mask material local differences.

If one focuses not on agreement with Kalven and Zeisel's data, but on agreement across the four sites, local variation also exists. Table 3 shows that the four sites had judge-jury agreement rates of 88 percent, 89 percent, 66 percent, and 64 percent—two higher agreement-rate jurisdictions and two lower agreement-rate jurisdictions. A pair-wise comparison of each of the sites indicates that the differences between higher and lower agreement-rate

---

[38]Hung Juries, supra note 10, at 6–9; Paula L. Hannaford, Valerie P. Hans & G. Thomas Munsterman, How Much Justice Hangs in the Balance? A New Look at Hung Jury Rates, 83 Judicature 59 (1999).

[39]For analysis of site-based differences of jurors' votes in the cases studied here, see Stephen P. Garvey, Paula Hannaford-Agor, Valerie P. Hans, Nicole L. Mott, G. Thomas Munsterman & Martin T. Wells, Juror First Votes in Criminal Trials, 1 J. Empirical Legal Stud. 371 (2004).

[40]Kalven & Zeisel, supra note 1, at 520–23.

Table 3: Judge-Jury Agreement, by Site

	Jury Acquits	*Jury Convicts*	*Jury Acquits*	*Jury Convicts*
	Los Angeles (88%; $p = 0.070$)		Maricopa (89%; $p = 0.508$)	
Judge acquits	7	1	14	6
Judge convicts	7	52	3	56
	Bronx (66%; $p = 0.093$)		DC (64%; $p < 0.001$)	
Judge acquits	9	7	7	2
Judge convicts	16	36	25	42

NOTE: NCSC data covering trials at four sites in 2000–2001. Cells report numbers of cases. Agreement rates and p values of disagreement pattern (which ignore cases of agreement) in parentheses.

sites are statistically significant ($p = 0.001$). That is, Los Angeles's rate of agreement significantly differs from both the Bronx's and DC's. Maricopa's rate of agreement similarly differs. But Los Angeles's agreement rate does not significantly differ from Maricopa's and the Bronx's does not significantly differ from DC's.

Interpretation of intersite disagreement patterns requires caution. Within-site variation between judge and jury is so revealing because the two adjudicators assess the same cases. Intersite comparisons require considering the fact that the sites process different cases. A lower agreement rate in the Bronx could be a function of different routing of cases to juries than the routing in, for example, Los Angeles. To illustrate, suppose Bronx juries have a reputation for being prone to acquit. That information would be factored into prosecution and defense decisions about which cases to plea bargain and which cases to press to trial. This pretrial process could lead to different mixes of cases reaching trial in the four jurisdictions. Interpreting the trial results in light of this selection process is as much art as science.[41] Our principal focus here is on inferences that can be drawn from judge-jury observation of the same cases.

2. Asymmetrical Disagreement

In Los Angeles, the Bronx, and DC, Table 3 shows that the asymmetry in judge-jury conviction tendency, when they see the same case, is in the direc-

[41]Cf. Kevin Clermont & Theodore Eisenberg, Litigation Realities, 88 Cornell L. Rev. 119, 140–42 (2002).

tion found by Kalven and Zeisel—judges tend to convict when juries would acquit more than juries tend to convict when judges would acquit. The asymmetry is statistically significant or nearly so in the three jurisdictions. In Maricopa, however, juries tend to convict when judges would convict, though the difference is not statistically significant. Again, it seems likely that Kalven and Zeisel's aggregate summaries mask interesting local variation.

III. Explaining Judge-Jury Disagreement

Section II indicates that judges and juries do sometimes disagree, and that the general direction of the disagreement suggests less judicial sympathy for defendants. This section explores some possible bases for judges' higher conviction rates: juries' different application of the beyond-a-reasonable-doubt standard, juries' different reaction to evidentiary complexity,[42] and juries' different reaction to legal complexity.

A. The Jury's Higher Conviction Threshold

Kalven and Zeisel, through a reasoned process of elimination, satisfied themselves that one source of judge-jury disagreement is that judges and juries apply different standards of proof beyond a reasonable doubt in criminal cases. "[T]he jury's more stringent view of proof beyond a reasonable doubt is one explanation for disagreements between the two deciders in criminal cases."[43] Kalven and Zeisel even regarded the major evidentiary effect they detected—the jury's greater tendency to believe the defendant—as subject to the competing explanation of disagreement: that the jury applied a higher evidentiary threshold to convict than the judge.[44]

The NCSC data contain evidence confirming this belief in the jury's more stringent view of proof beyond a reasonable doubt. However, the data suggest a pattern somewhat more complex than Kalven and Zeisel were able to report.

---

[42]For an analysis of complexity as perceived by judges, juries, and attorneys, see Michael Heise, Criminal Case Complexity: An Empirical Perspective, 1 J. Empirical Legal Stud. 331 (2004).

[43]Kalven & Zeisel, supra note 1, at 189.

[44]Id. at 181.

Table 4: Judge and Jury Agreement, by Evidentiary Strength of Case

	Adjudicator Whose View of Evidentiary Strength Classifies Evidence as Weak, Medium, or Strong			
	Judge		*Jury*	
	Jury Acquits	*Jury Convicts*	*Jury Acquits*	*Jury Convicts*
*A. Cases in Which Evidence of Conviction is **WEAK** (1 or 2 on Likert Scale)*				
Judge acquits	14	5	15	2
Judge convicts	0	7	9	3
*B. Cases in Which Evidence of Conviction is **MEDIUM** (3, 4, or 5 on Likert Scale)*				
Judge acquits	23	10	21	10
Judge convicts	35	81	37	81
*C. Cases in Which Evidence of Conviction is **STRONG** (6 or 7 on Likert Scale)*				
Judge acquits	0	1	0	3
Judge convicts	15	97	2	92

NOTE: NCSC data covering trials at four sites in 2000–2001. Cells report numbers of cases.

NCSC questionnaires asked both judges and juries to rank, on a scale of one to seven, the strength of the evidence favoring conviction, with 1 representing the weakest evidence of conviction and 7 the strongest. Panels A, B, and C of Table 4 report judge-jury agreement data for each of three levels of evidentiary strength: weak (1 and 2 on the scale), medium (3, 4, and 5 on the scale), and strong (6 and 7 on the scale). Table 4 reports in its "Judge" columns the rates of agreement with the cases classified by the *judge's* view of the strength of the evidence. Table 4's "Jury" columns report judge-jury agreement with the cases classified by the *jury's* view of the strength of the evidence. The jury values are rounded based on the average of the individual jurors' responses.

For example, Panel A reports results for cases in which the evidence favoring conviction was weak. According to the "Judge" columns, when the judge regarded the evidence as weak, the judge and jury agreed on acquittal in 14 cases and agreed on conviction in seven cases. The jury would convict when the judge would not in five cases and the judge would convict when the jury would not in zero cases. Thus, not surprisingly, when the judge regarded the evidence as weak, the judge tended to convict less than the jury. Panel A's "Jury" columns show the opposite effect when the evidentiary strength is classified as weak based on the jurors' reports of the evidence. That is, when the jury regarded the evidence as weak, the jury would acquit

in nine cases in which the judge would convict, and the judge would acquit in only two cases in which the jury would convict.

Panel A suggests the importance of whose view of the evidentiary strength is used. Kalven and Zeisel had only the judges' views and were criticized on that ground.[45] A richer and more complex picture emerges by recognizing that evidentiary strength may be perceived differently by different observers.

Panel A's cases comprise a small fraction of the total because neither the judge nor the jury often regard the evidence as weak. Prosecutors do not randomly press cases to trial and, on average, the evidence should be well above the subjective level of "weak." In Panel B, where we classify the evidence as being of medium strength, an important pattern emerges in a larger class of cases. Here, regardless of which observer classifies evidentiary strength, the judge is more prone to convict than the jury. Judges who regard the evidentiary strength as medium would convict in 35 cases in which the jury acquitted, and the jury convicted in only 10 cases in which the judge would acquit. When the jury regards the evidentiary strength as medium, the judge would convict in 37 cases in which the jury acquitted, and the jury convicted in only 10 cases in which the judge would acquit.

Panel C, which also covers many more cases than Panel A, again shows the importance of whose evidentiary standards apply. Judges who regard the evidence favoring conviction as strong were willing to convict 15 defendants that the jury acquitted. When the jury regards the evidence as strong, the judge would convict in all but three cases.

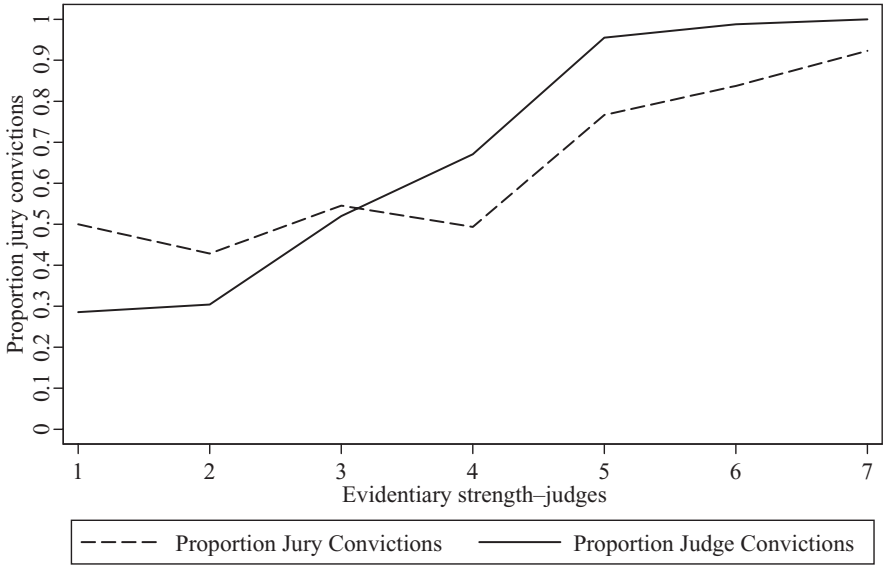
Graphical representation of the judge and jury conviction behavior as a function of evidentiary strength is also revealing. Figure 1 shows the relation between the *judges'* views of the evidence and the rate at which judges would, and juries did, convict. For each of the seven levels of evidentiary strength, the figure reports the mean conviction rate for juries in one line and for judges in the other. The figure suggests that the jury had more of a "yes-no" evidentiary threshold for conviction than did the judge. The direction of this jury tendency is clear. Juries required a case more favorable to the prosecution to support a conviction.

Figure 1 shows that, unless the judges' views of the evidence favoring the prosecution reached the level of at least "5" on the 1–7 scale, juries did not consistently vote to convict in more than 50 percent of the cases. Judges,

---

[45]Goldstein, supra note 8.

Figure 1: Convictions and evidentiary strength: judges' view of the evidence.

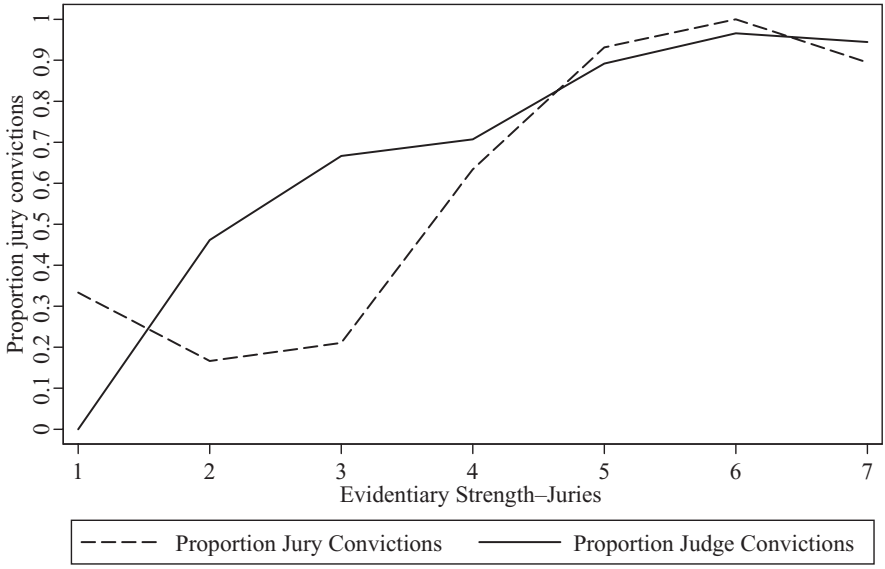


NOTE: NCSC data covering trials at four sites in 2000–2001. Each line shows on the y-axis the mean conviction rate at each of seven levels of evidentiary strength, as evaluated by judges. Higher numbers along the x-axis indicate higher levels of evidentiary strength favoring conviction. The "Proportion Jury Convictions" line shows the observed conviction rates. The "Proportion Judge Convictions" line is based on the judges' indication whether they would convict.
SOURCE: NCSC data covering trials at four sites in 2000–2001.

in contrast, would have exceeded a 50 percent conviction rate for evidence reaching the level of a "4" on the 1–7 scale. The jump in the jury conviction rate from cases with evidentiary strength reported by the judge as "4" to cases with evidence strength reported by the judge as "5" is striking. Jurors convicted at about a 50 percent rate at the "4" (or lower) level. Their conviction rate jumps to near 80 percent for evidence at the "5" or higher level. Juries required that a case be at least a "5" on the judge's closeness-of-case scale to support high conviction rates. Nearly half the juries found reasonable doubt below what the judge regarded as a strong evidentiary showing by the prosecutor.

Figure 1's second line shows that judges also exhibit a jump in willingness to convict as evidence crosses from the equivocal "4" to the pro-prosecution "5." But the judicial jump is from the relatively higher conviction rate of 65 percent to over 90 percent.

Figure 2: Convictions and evidentiary strength: juries' view of the evidence.



SOURCE: NCSC data covering trials at four sites in 2000–2001.

Employing the *juries'* view of the closeness of the evidence, rather than the judges', reinforces Figure 1's key findings. Figure 2 uses the juries' reports of how much the evidence favored the prosecution, also on a 1–7 scale, with the jury average computed from the jurors' individual responses and then rounded to an integer. Judges are willing to convict in cases much less favorable to the prosecution, as ranked by the juries' view of the evidence, than are juries. Quantum increases in juries' willingness to convict emerge as the evidence favoring the prosecution moves from "3" to "4" and from "4" to "5." Judges' willingness to convict exceeds 50 percent of cases even when jurors regard the evidence as largely favorable to the defense—"3" on the 1–7 scale. As the evidence favoring the prosecution moves from "4" to "5," jurors' willingness to convict soars to over 90 percent. Note also that the judges' reported rate of willingness to convict is almost monotonically increasing as the jurors regard the evidence as increasingly pro-prosecution. Judges and juries seem to be reacting to the evidence in a similar manner, except that juries require stronger evidence to convict than judges do.

Table 5: Judge and Jury Conviction Rates, by Evidentiary Complexity

	Adjudicator Whose View of Evidence Complexity Classifies Complexity as Low, Medium, or High			
	Judge		*Jury*	
	Jury Acquits	*Jury Convicts*	*Jury Acquits*	*Jury Convicts*
*A. Cases in Which Complexity is **LOW** (1 or 2 on Likert Scale)*				
Judge acquit	22	11	11	3
Judge convict	30	109	4	27
*B. Cases in Which Complexity is **MEDIUM** (3, 4, or 5 on Likert Scale)*				
Judge acquit	14	4	25	11
Judge convict	16	70	42	136
*C. Cases in Which Complexity is **HIGH** (6 or 7 on Likert Scale)*				
Judge acquit	1	1	0	1
Judge convict	5	6	2	13

NOTE: NCSC data covering trials at four sites in 2000–2001. Cells report numbers of cases.

B. Evidentiary Complexity as an Explanation of Disagreement

Kalven and Zeisel concluded that juries understood the evidence well and that juror confusion about the evidence was not the source of much disagreement. The above discussion of evidentiary strength suggests the importance of considering which decider's view of the evidence to use. Similarly, complexity is not a purely objective concept. Judges and juries might well regard evidentiary complexity differently.[46] So it is important to consider complexity's influence on agreement rates from both the judges' and the juries' perspectives. The NCSC questionnaires asked both judges and jurors to rank each case's evidentiary complexity on a one to seven Likert scale. We can assess complexity's influence on agreement using both the judges' and the juries' reports of complexity.

Table 5 mimics Table 4's structure by showing agreement rates separately for judge and jury reports of complexity (in contrast to the evidentiary strength that was Table 4's focus). Preliminarily, the table shows that judges tended to regard most cases as being of low complexity whereas jurors tended to regard most cases as being of medium complexity. Neither regarded even 10 percent of the cases as being highly complex.

[46]Heise, supra note 42.

The table also suggests that rates of disagreement are not strongly associated with complexity, regardless of which adjudicator's assessment of complexity is used. In cases in which the judge regarded the evidence as being of low complexity, the adjudicators agreed on the outcome in 131 of 172 (76 percent) cases. When the judge regarded the evidence as being of medium complexity, they agreed on 84 of 104 (81 percent) of the cases. In cases in which the jury regarded the evidence as being of low complexity, the adjudicators agreed on the outcome in 38 of 45 (84 percent) cases. When the jury regarded the evidence as being of medium complexity, they agreed on 161 of 214 (75 percent) of the cases, not significantly different than the agreement rate in low complexity cases ($p = 0.243$). The strongest hint of increasing disagreement with increasing complexity is the agreement rate in seven of the 13 cases (54 percent) that the judge regarded as being of high complexity.

Table 5 also shows the persistent asymmetry in cases of judge-jury disagreement. In both Panels A and B, judges would convict when juries acquitted more than juries convicted when judges would acquit. The judge's lower conviction threshold does not appear to be a function of case complexity, regardless of which decider's view of complexity is used.

In summary, the data support Kalven and Zeisel's conclusion that perceived complexity is not a promising explanation of judge-jury disagreement. The data are more consistent with the jury having a higher threshold for conviction.

C. Legal Complexity as an Explanation of Disagreement

The NCSC questionnaire also asked the judges to rank the complexity of the law in each case on a one to seven scale. Understandably, the NCSC did not ask the jurors to opine about the question of legal complexity. To analyze the judges' views as they may relate to judge-jury disagreement, we again divide the sample into cases of low, medium, and high complexity. Table 6 reports the disagreement pattern separately for each of the three legal complexity levels.

For cases ranked low in legal complexity, judges and juries agreed in 116 of 149 (78 percent) cases. At the medium level of complexity, they agreed in 85 of 111 (77 percent) of cases. In cases of high legal complexity, they agreed in 21 of 29 (72 percent) cases, a rate insignificantly different from the rate in low complexity cases ($p = 0.630$). Increasing legal complexity thus is also not a promising explanation of judge-jury disagreement. Table

Table 6: Judge and Jury Conviction Rates, by Judges' View of Cases' Legal Complexity

	Jury Acquits	*Jury Convicts*
*A. Cases in Which Legal Complexity is **LOW** (1 or 2 on Likert Scale)*		
Judge acquit	22	7
Judge convict	26	94
*B. Cases in Which Legal Complexity is **MEDIUM** (3, 4, or 5 on Likert Scale)*		
Judge acquit	12	8
Judge convict	18	73
*C. Cases in Which Legal Complexity is **HIGH** (6 or 7 on Likert Scale)*		
Judge acquit	3	1
Judge convict	7	18

NOTE: NCSC data covering trials at four sites in 2000–2001. Cells report numbers of cases.

6 also shows that the asymmetry of disagreement survives at each level of complexity.

In sum, neither legal complexity nor evidentiary complexity is helpful in explaining the overall pattern of judge-jury disagreement.

D. Difference in Agreement Rates Across Crimes

Kalven and Zeisel report judge-jury agreement rates for each of 42 crimes.[47] The NCSC data identify 19 different crimes; a useful summary of the major crimes by locale appears elsewhere.[48] Based on analyses not reported here, little insight is gained into the pattern of disagreement by focusing on individual crimes, with two exceptions. In two crime categories, site-specific rates of disagreement are worth noting. Drug-sale cases in the Bronx produce a judge-jury agreement rate of only 53.8 percent in 13 cases with usable data. Weapons cases in the District of Columbia produce agreement in only 33.3 percent of 12 cases. The small numbers in the site-specific drug sales and weapons cases suggest caution in assigning substantial importance to these disagreement rates. We defer incorporating this information into regression models analysis until considering other factors.

---

[47]Kalven & Zeisel, supra note 1, at 66–81.

[48]Hung Juries, supra note 10, at 36.

E. Juror Characteristics

Substantial evidence exists both questioning and supporting the influence of juror demographic characteristics on case outcomes.[49] The NCSC data include responses from more than 3,000 jurors. But Kalven and Zeisel reported their results at the jury level. So we do not address here the relation between individual juror votes and case outcomes.[50] This analysis studies the influence of juror demographic factors by aggregating juror characteristics at the case level.[51] For example, if six of ten juror questionnaires in a case indicate that the juror was male, the variable "male" would have the value 0.6 in the jury-level aggregated data. Table 7, Panel A, provides descriptive statistics for juries.

Panel B reports jury characteristics broken down by patterns of agreement with judges. For example, Panel B's second row shows that the percent of African Americans on juries was fairly constant, 20 percent to 23 percent, for cases in which judges and juries agreed about conviction or acquittal, and in which the judge would acquit where the jury convicted. In the small class of cases in which the jury acquitted when the judge would have convicted, the mean percent of African Americans on juries was substantially higher, 37 percent. Similarly, larger Hispanic presence on the jury appears to be associated with increased likelihood of the jury convicting when the judge would acquit.

We again defer formal tests of these effects until we combine these variables with evidentiary strength variables in regression models.

IV. Regression Analysis

Section III suggests that a different conviction threshold, and not evidentiary or legal complexity, is a strong candidate for explaining judge-jury dif-

---

[49]Note 60 intra.

[50]For analysis of the relation between juror characteristics and their votes in these data, see Garvey et al., supra note 39.

[51]Such aggregation can mask juror demographic tendencies. Theodore Eisenberg, Stephen P. Garvey & Martin T. Wells, Forecasting Life and Death: Juror Race, Religion, and Attitude Toward the Death Penalty, 30 J. Legal Stud. 277, 302 (2001) ("race and religion play a statistically significant role in all of the first-vote models, but in all the final-vote models their significance becomes marginal or dissolves altogether") (footnote omitted).

Table 7: Judge-Jury Agreement and Juror Characteristics

A. Descriptive Statistics of Jury Characteristics

	Mean	*Median*	*N*
Percent male jurors	42%	44%	361
Percent African-American jurors	26%	20%	363
Percent Hispanic jurors	22%	17%	363
Percent previously served on jury	53%	50%	364
Education (1–5 scale)	3.49	3.50	364

B. Juror Characteristics and Judge-Jury Agreement

	Judge Convict		*Judge Acquit*	
	Jury Convict	*Jury Acquit*	*Jury Acquit*	*Jury Convict*
Percent male jurors	44%	38%	37%	42%
Percent African-American jurors	23%	37%	20%	21%
Percent Hispanic jurors	22%	18%	24%	29%
Percent previously served on jury	52%	54%	51%	49%
Education (1–5 scale)	3.46	3.62	3.57	3.18

NOTE: Panel B percents are the percent of jurors with the row characteristic comprising the jurors who sat in cases with the column characteristic. For example, males comprised 44 percent of the jurors who sat in cases in which the judge would convict and the jury did convict. The table is based on the NCSC data covering four sites in 2000–2001. Education coding: 1 = not high school graduate, 2 = high school graduate, 3 = some college, 4 = college graduate, 5 = postgraduate work.

ferences. However, a different threshold can hardly be the sole factor worth considering. Kalven and Zeisel explored several other factors. Shifting from univariate to multivariate analysis, our goal is to explore whether Section III's core result—the different conviction threshold—survives accounting for other salient characteristics.

A. The Conviction Threshold

Logistic regression models allow examination of conviction as a function of multiple factors, including strength of evidence. The dependent variable in the models is conviction. To allow simultaneous consideration of both adjudicators' conviction propensity, as well as other factors, the data are organized as two observations of each trial, one observation by the judge and one observation by the jury. The dependent variable "convict" is coded as 1 for the jury observation if the jury did convict and for the judge observation if

the judge would have convicted. It is coded 0 for the jury observation if the jury did not convict and for the judge observation if the judge would not have convicted. To capture the adjudicator effect, the models include a judge dummy variable. It is coded 1 for all the judicial observations of the data.

Since each case is "observed" twice, four possible combinations of values of the "convict" variable and the judge dummy variable exist. If judge and jury agree on conviction, "convict" is 1 for both values (0 and 1) of the judge dummy variable. If the judge would convict but the jury acquitted, the judge observation of the trial would have "convict" equal to 1 and the jury observation of the trial (for which the judge dummy variable is 0) would have "convict" equal to 0. If the jury convicted but the judge would have acquitted in the same case, the judge observation of the trial would have "convict" equal to 0 and the jury observation of the trial (for which the judge dummy variable is 0) would have "convict" equal to 1. If judge and jury agree on acquittal, "convict" is 0 for both values (0 and 1) of the judge dummy variable. Table 8 enumerates the possible coding combinations of the key variables—"Convict" and "Judge Dummy"—in four hypothetical trials.

To explore whether judge and jury conviction thresholds differ, while holding other factors constant, the variable of key interest is the judge dummy explanatory variable. Except for the judge dummy variable, the explanatory variables for the two observations of each case are identical. The dependent variable, "Convict," can also vary across the two observers of each case, judge and jury.

Table 8: Coding of Conviction and Judge Dummy Variables

Trial Number	*Dependent Variable = "Convict"*	*Judge Dummy Explanatory Variable*	*Other Explanatory Variables*	*Verbal Description*
1	1	1	Constant within Trial 1	Judge would convict; jury did
	1	0		
2	1	1	Constant within Trial 2	Judge would convict; jury didn't
	0	0		
3	0	1	Constant within Trial 3	Judge would acquit; jury did
	1	0		
4	0	1	Constant within Trial 4	Judge would acquit; jury didn't
	0	0		

Kalven and Zeisel's findings, and Sections II and III above, forecast that the judge dummy variable's coefficient should be positive—corresponding to an increased likelihood of conviction compared to the jury. That is, the likelihood of conviction should increase with the observation of the trial being by the judge rather than by the jury. To test the robustness of this result, and to explore other possible influences on conviction, we model the dependent variable "convict" as a function of the judge dummy variable, of the juror characteristics described in Table 7, of the crime characteristics described in Section III.D above, and of the site of the trial.

Table 9 reports the results. Models 1–4 account for the stratification of the sample by site. Models 5–8 employ dummy variables for each site, with Los Angeles serving as the reference category. For each of the two groups of models, three measures of the evidentiary strength favoring conviction are used. Models 1 and 5 include both the judges' and the juries' reports of evidentiary strength. Models 2 and 6 include only the judges' reports of evidentiary strength. Models 3 and 7 include only the juries' reports of evidentiary strength. Models 4 and 8 employ no measure of evidentiary strength.

1. Judge Effects

In all models the judge dummy variable is positive and statistically significant. Regardless of whose view of evidentiary strength is used, judges tend to convict more than juries. Controls for juror characteristics, case characteristics, and locale do not change this core result. The magnitude of the effect is substantial. Depending on the model selected, the marginal effect of a judge trial compared to a jury trial is at least a 12 percent increase in probability of conviction.

2. Evidentiary Strength Effects

Table 9 yields other interesting results. Studies suggest that evidentiary strength so dominates case outcomes that influences of other factors are difficult to detect.[52] The data offer a rare nonexperimental opportunity to test the influence of other factors while controlling for the strength of the evidence. The evidentiary strength variables behave as one would expect. In all models in which they are used, all evidentiary strength variable coefficients

---

[52]Theodore Eisenberg & Martin T. Wells, Trial Outcomes and Demographics: Is There a Bronx Effect? 80 Texas L. Rev. 1850 & n.48 (2002) (summarizing studies).

Table 9: Logistic Regression Models of Conviction (Dependent Variable = Convict)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Judge	1.130***	0.836***	1.037***	0.669***	1.140***	0.838***	1.046***	0.681***
	(3.99)	(4.23)	(4.00)	(4.22)	(3.92)	(4.12)	(3.93)	(4.10)
Evidentiary strength—juries	1.108***		1.225***		1.135***		1.246***	
	(8.43)		(9.84)		(8.13)		(9.35)	
Evidentiary strength—judges	0.622***	0.806***			0.624***	0.795***		
	(6.54)	(8.59)			(6.38)	(8.28)		
Male percent of jury	1.124	1.335**	1.012	1.112*	1.148	1.237**	1.040	0.860
	(1.44)	(2.19)	(1.37)	(1.92)	(1.45)	(2.04)	(1.41)	(1.52)
African-American percent of jury	1.479**	0.300	1.306**	–0.037	1.075	0.306	0.822	0.102
	(2.25)	(0.54)	(2.07)	(0.06)	(1.21)	(0.44)	(1.02)	(0.14)
Hispanic percent of jury	0.354	0.189	–0.003	–0.271	–0.580	0.028	–0.883	–0.038
	(0.45)	(0.27)	(0.00)	(0.39)	(0.60)	(0.03)	(0.92)	(0.05)
Education level (1–5 scale)	–0.466*	–0.371	–0.413	–0.187	–0.327	–0.408	–0.387	–0.432
	(1.95)	(1.57)	(1.62)	(0.81)	(0.98)	(1.38)	(1.25)	(1.54)
Weapons charge dummy	–0.291	–1.268***	–0.166	–1.195***	–0.243	–1.225***	–0.120	–1.166***
	(0.63)	(3.17)	(0.38)	(3.24)	(0.51)	(2.98)	(0.27)	(3.10)
Drug-possession charge dummy	0.069	–0.297	0.148	–0.176	0.004	–0.390	0.021	–0.313
	(0.17)	(0.58)	(0.32)	(0.40)	(0.01)	(0.72)	(0.04)	(0.68)
Maricopa					–0.617	–0.402	–0.846*	–0.492
					(1.26)	(0.85)	(1.75)	(1.16)
Bronx					0.188	–0.418	–0.097	–0.995**
					(0.41)	(0.86)	(0.21)	(2.27)
DC					–0.505	–0.361	–0.507	–0.261
					(0.85)	(0.79)	(0.87)	(0.56)
Constant	–5.909***	–2.093*	–3.614***	1.195	–5.961***	–1.524	–3.102**	2.512**
	(4.45)	(1.90)	(2.86)	(1.15)	(3.55)	(1.11)	(2.20)	(2.13)
Observations	540	540	544	544	540	540	544	544

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

NOTES: Absolute value of t statistics in parentheses, adjusted for multiple observations per case. NCSC data covering trials at four sites in 2000–2001. The number of cases is half the number of observations because each observation is duplicated to allow for a judge and jury outcome for each case. Los Angeles is the reference category for the locale dummy variables in Models 5–8.

are positive and highly statistically significant. Regardless of which adjudicator's view of evidentiary strength is used, stronger evidence of guilt is strongly associated with increased likelihood of conviction.

One concern about using evidentiary strength as an explanatory variable is that assessments of evidentiary strength in a case may not be independent of juror characteristics or even of case characteristics. Male jurors may react to evidence differently than female jurors. Jurors may react to evidence in a weapons case, where no one is harmed and no property is taken, differently than they react to evidence in other cases. The evidentiary strength variables may be absorbing some of what is interesting in the other variables. Models 4 and 8 address this concern by omitting both evidentiary strength variables. Their results are reasonably consistent with the models that include evidentiary strength variables.

3. Case and Juror Effects

Holding constant the evidence, Table 9 offers little robust support for the conventional wisdom that an increased African-American presence on juries corresponds to a lower likelihood of conviction.[53] The coefficient on the "African-American percent of jury" variable is positive in seven of eight models and is statistically significant in two models, in the direction of greater conviction likelihood. The coefficient changes sign in one model. Similarly, the Hispanic percent of juries provides little explanatory power about case outcomes.[54]

Table 9 provides some evidence that the male percentage of the jury is associated with increased conviction likelihood. In all eight models, the "Male percent of jury" coefficient is positive and in three it is statistically sig-

---

[53]Since we study only aggregated juror effects, the finding of no race effect does not eliminate the possibility that race effects would emerge in studying individual juror votes. See Eisenberg et al., supra note 51, at 300–01; Theodore Eisenberg, Stephen P. Garvey & Martin T. Wells, Victim Characteristics and Victim Impact Evidence in South Carolina Capital Cases, 88 Cornell L. Rev. 306, 334 tbl. 7 (2003).

[54]The significance of the "African-American percent of jury" variable in Models 1 and 3 is in part a function of a significant relation between jurors' view of evidentiary strength and race. In models of juror view of evidentiary strength, African-American jurors view evidence as less strong than white jurors. However, this effect does not persist when one controls for site. To fully explore the relation between juror characteristics and conviction, one could develop models that account for possible endogeneity in the relation between evidentiary strength and juror characteristics.

nificant or nearly so. Similarly, increased education levels correspond with a decreased likelihood of conviction in all models. The "Education level" coefficient is negative and of reasonably stable magnitude in all models, but is nearly statistically significant in only one model.[55]

Table 9 also shows a consistently negative, and sometimes significant, association between the trial being about a weapons charge and the likelihood of conviction. Perhaps weapons charges often are a last-gasp count charged by prosecutors who lack sufficient evidence of a murder, robbery, or attempt, but wish to incarcerate someone they regard as a bad actor. Adjudicators may react negatively to what appears to be close to an anticipatory prosecution. In addition, jurisdictions often have mandatory sentences or enhancements in cases involving weapons offenses.[56] If juries know of such provisions, and regard mandatory sentences or enhancements for such crimes as unfair, one may observe lower conviction rates in such cases.

Other plausible models of conviction are worth exploring. If the higher-threshold theory is correct, the judges' greater conviction proneness ought to emerge when evidence is viewed from the perspective of third parties. The NCSC data contain not only the judges' and juries' reports of evidentiary strength, but also those of prosecutors and defense attorneys. We have explored models similar to those in Table 9 that use these others' views of evidentiary strength to control for the merits of cases. In virtually all such models, the judge dummy variable is associated with increased likelihood of conviction. The one exception, echoing Table 4's results, is that when one uses the judges' views of the evidence, and the view is that the evidence is weak, the judge is less likely to convict than the jury. Table 9's results suggest that this effect is swamped by the many cases of middling or strong evidentiary strength, in which the judge is uniformly more likely to convict than the jury.

---

[55]Endogeneity, see note 54 supra, is less of a concern with respect to "Male percent of jury" and "Education level" because these variables are not statistically significant in models of evidentiary strength.

[56]For example, Ariz. Rev. Stat. §13–604(F) (2001) (person convicted of lesser felonies involving the use or threatening exhibition of a deadly weapon "shall not be eligible for suspension of sentence, probation, pardon or release from confinement on any basis" except narrow conditions); Cal. Penal Code §§12020(a)(1) (up to one-year prison term for possession of many weapons); 1203(e)(11) (limiting probation chances for certain weapons-related offenses).

4. Site-Level Analysis

Table 9 aggregates the data, yet Table 3 indicates that local variation exists. The Appendix reports models similar to those in Table 9 for each of the four sites separately. The models confirm the pattern suggested by Table 3's site-level 2×2 tables. In Los Angeles and DC, judges are significantly more likely to convict than juries. In the Bronx, the effect is in the same direction but is not statistically significant. In Maricopa, the effect is in the opposite direction but is not statistically significant. Overall, the data support the judges' greater conviction proneness but also show that the effect is not universal and should not be assumed to exist in every locale.

The Appendix also indicates that the relation between conviction and juror characteristics is not constant across sites. The coefficient for "Male percent of jury" is positive for two sites and negative for two sites. The coefficients on the African-American and Hispanic percentages of jurors also change signs across sites. Increased education is associated with lower conviction rates in three of four sites and is statistically significant or nearly significant in two of them. Overall, the education effect is the most robust across the aggregate models in Table 9 and the site-level models. But even this effect is not universal.

B. Explaining Judge-Jury Differences

Table 9 suggests more rigorously than heretofore established the greater conviction proneness of judges, at least for our sample. It does not reveal the cases in which judges tend to disagree with juries or explain the cases in which they disagree. Models of conviction, standing alone, cannot reveal the pattern of disagreement.

Yet it is the pattern of disagreement that most interested Kalven and Zeisel. They limited much of their book to discussing cases in which judges and juries disagreed. After noting the generally high rates of judge-jury agreement, they all but excluded the agreement cases from their analysis. They classified disagreements as relating to guilt, charge, or hung-jury status, and then stated: "The precise quest of this study then is to explain what caused the disagreements in these 1063 instances, constituting 30 percent of all trials."[57]

---

[57]Kalven & Zeisel, supra note 1, at 110.

Table 7, Panel B, reports summary statistics for jury characteristics across the possible combination of judge-jury agreement and disagreement. To further explore disagreement, we divide the cases into three categories—those in which the judge and jury agreed, those in which the judge would have convicted but the jury acquitted, and those in which the jury convicted but the judge would have acquitted. These categories, whose occurrence rates are reported in Table 1, Panel A, are used to construct a categorical dependent variable in multinomial logit models.[58] The models employ the same sets of explanatory variables as are used in Table 9.

Table 10 reports the results. Models 1–3 account for the stratified structure of the sample by site. Models 4–6 use dummy variables to control for locale effects. Table 10 suggests that education, a factor helpful in explaining jurors' conviction proneness in Table 9, also helps explain judge-jury disagreement. In all models, the coefficient for "Education level" for the jury-convicts/judge-acquits outcome, compared to judge-jury agree outcome, is negative. This suggests that increased education is associated with the jury not convicting compared to agreeing with the judge. The coefficient for education for the judge-convicts outcome is consistently positive. This suggests that greater jury education is associated with the judge convicting when the jury would not. In all models other than Model 6, the two education coefficients are statistically significantly different from each other ($p < 0.05$) or nearly so ($p < 0.10$), even when not statistically significantly different from the agreement reference category.

The two race variables, which played inconsistent roles in explaining the tendency to convict, are more helpful in explaining judge-jury differences, but the evidence is mixed. In Models 1–3, an increasing African-American percentage of jurors strongly correlates with the judge tending to convict when the jury would acquit, but the effect diminishes and becomes insignificant in Models 4–6, those with locale dummy variables. In Models 4–6, increased Hispanic percentage of the jury is associated with an increased tendency of the jury to convict when the judge would acquit. To the extent we find evidence of minority effects, they are more complex than conventional wisdom suggests.

Site-specific models similar to those in Table 10 are difficult to estimate because of the small number of cases fitting into one or the other of the two disagreement categories. Table 3 shows that only the Bronx has more than

[58]One could also plausibly regard the dependent variable as ordered, but richer information emerges from a multinomial logit model.

Table 10: Multinomial Logit Models of Agreement (Dependent Variable = Jury Convict/Judge Acquit, Judge Convict/Jury Acquit, Agree (Reference Category))

	Adjudicator Willing to Convict When Other Would Acquit											
	Jury	*Judge*	*Jury*	*Judge*	*Jury*	*Judge*	*Jury*	*Judge*	*Jury*	*Judge*	*Jury*	*Judge*
	1		*2*		*3*		*4*		*5*		*6*	
Male	–0.355	–0.385	0.067	–0.527	–0.299	0.038	0.045	–0.105	0.230	–0.206	0.098	0.158
percent of jury	–0.27	–0.43	–0.05	–0.05	–0.23	–0.04	–0.04	–0.13	–0.19	–0.25	–0.08	–0.16
African-American	–1.120	2.303***	–1.306	2.406***	–1.219	1.932**	–1.690	0.522	–1.44	0.703	–1.900	0.206
percent of jury	–0.90	–3.16	–0.98	–3.21	–0.97	–2.37	–1.20	–0.56	–0.93	–0.74	–1.25	–0.18
Hispanic	0.490	–0.186	0.028	–0.346	0.508	–0.660	0.261	–1.422	0.328	–1.600	0.285	–1.409
percent of jury	–0.45	–0.22	–0.03	–0.39	–0.44	–0.60	–0.23	–1.47	–0.24	–1.64	–0.24	–1.03
Education	–1.047*	0.584*	–0.967	0.617*	–1.051*	0.663*	–0.591	0.422	–0.692	0.506	–0.687	0.199
level (1–5 scale)	–1.87	–1.70	–1.51	–1.76	–1.86	–1.73	–1.22	–1.09	–1.28	–1.31	–1.35	–0.38
Weapons	1.640*	1.170**	1.714*	1.168**	1.446	0.359	1.422	0.988*	1.557	0.988*	1.227	0.092
charge dummy	–1.86	–2.16	–1.83	–2.14	–1.58	–0.54	–1.53	–1.80	–1.50	–1.78	–1.31	–0.14

Drug-possession charge dummy	−0.343	−0.548	−0.349	−0.510	−0.406	−1.087	−0.068	−0.358	−0.300	−0.310	−0.104	−1.092
	−0.32	−0.69	−0.34	−0.63	−0.37	−1.24	−0.06	−0.46	−0.30	−0.39	−0.09	−1.26
Evidentiary strength—judges			−0.688***	0.014					−0.682***	0.025		
			−3.82	−0.13					−3.36	−0.23		
Evidentiary strength—juries					−0.178	−0.882***					−0.141	−0.935***
					−1.04	−5.37					−0.85	−5.31
Maricopa							1.325	−1.173	1.335	−1.022	1.297	−1.573*
							−1.10	−1.61	−0.92	−1.37	−1.07	−1.84
Bronx							2.170**	0.981	1.566	1.137*	2.074*	0.036
							−2.01	−1.56	−1.51	−1.72	−1.87	−0.05
DC							1.338	0.539	1.306	0.571	1.458	0.614
							−1.00	−0.96	−0.86	−0.95	−1.11	−0.86
Constant	1.023	−4.122***	3.549	−4.285**	1.878	−0.729	−1.879	−3.140*	1.296	−3.653**	−0.845	1.749
	−0.51	−2.65	−1.45	−2.53	−0.86	−0.46	−0.94	−1.91	−0.60	−2.00	−0.35	−0.77
Observations	272		270		272		272		270		272	

*Coefficients, relative to reference category, significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

NOTES: Absolute value of t statistics in parentheses, adjusted for multiple observations per case. NCSC data covering trials at four sites in 2000–2001. Los Angeles is the reference category for the locale dummy variables.

three cases in each of the two disagreement categories. The other sites' models generate unrealistically high coefficients and standard errors because of the lack of variation in outcomes. In a Bronx-only model using only juror characteristics as explanatory variables, only education statistically significantly differs between the judge-would-convict and the jury-would-convict outcomes. As in Table 10, increased education corresponds to the judge convicting when the jury would not.

V. Conclusion

By controlling for multiple observers' views of evidentiary strength, we can confirm with additional rigor, albeit in a smaller sample, Kalven and Zeisel's finding that judges tend to convict more than juries—at least in the class of cases selected for trial by jury. We find little evidence that this effect is a function of evidentiary complexity or legal complexity. Judges simply appear to have a higher conviction threshold than juries. But we do not find evidence that this effect persists in every locale. A replication with more locales is needed to fully explore the persistence of the different conviction threshold.

Juror effects include a greater willingness of male jurors to convict, and more highly educated juries being less willing to convict than judges. Minority juror effects are mixed. They provide little explanatory power of convictions and somewhat greater power in explaining when judges and juries disagree. However, the effects do not persist at significant levels in models that control for locale. A richer set of locales is necessary to sort out minority-group effects, again suggesting the need for a larger study.

To the extent the NCSC data suggest substantial judge-jury agreement, the results may be conservative. The NCSC sample of districts includes three sites, Los Angeles, the Bronx, and DC, with substantial or dominant minority populations. Some believe that convictions are more difficult to obtain in places with high minority populations, such as the Bronx.[59] Suppose one accepts that minorities are less conviction prone than whites.[60] The racial

---

[59]Arthur S. Hayes, Inner City Jurors Tend to Rebuff Prosecutors and to Back Plaintiffs, Wall St. J. Mar. 24, 1994, p. 1, col. 1.

[60]For references to the literature on stereotypical views of jurors, see Eisenberg & Wells, supra note 52, at 1839–40 n.3, 1842–43 & nn.17–18 (2002). In general, researchers are skeptical about

breakdown of judges in the four sites strongly suggests that judges in these jurisdictions are less likely to be minorities than are jurors.[61] In the Bronx, for example, 21 of 29 judges (72 percent) are white but only 14 percent of the jurors who sat in sampled cases were white Caucasians.[62] In Los Angeles, 41 of 52 (79 percent) judges hearing criminal cases are white but only 36 percent of the jurors who sat in sampled cases were white Caucasians. If these sharp differences in racial composition of judge and juror groups promote disagreement, then even stronger agreement might be found at sites in which judge and jury racial proportions are more alike.

---

the existence of strong juror demographic effects on case outcomes. Id. at 1850 & n.48. However, there is growing evidence of juror race effects in capital sentencing. William J. Bowers et al., Death Sentencing in Black and White: An Empirical Analysis of the Role of Jurors' Race and Jury Racial Composition, 3 U. Pa. J. Const. L. 171 (2001) (African-American jurors more reluctant than white jurors to vote for death in capital cases); Eisenberg et al., supra note 51 (African-American jurors more reluctant than white jurors to vote for death in capital cases); Garvey et al., supra note 39; Samuel R. Sommers & Phoebe C. Ellsworth, How Much Do We Really Know about Race and Juries? A Review of Social Science Theory and Research, 78 Chi.-Kent L. Rev. 997 (2003). For other results consistent with race effects, see Joan M. Cheever & Joanne Naimann, The View from the Jury Box, Nat'l L.J., Feb. 22, 1993, p. S2 ("Black and white jurors had starkly contrasting views on how race influences the results of both civil and criminal trials, with blacks perceiving the system heavily weighted against minorities."); The View from the Jury Box: Racial Divide Affects Black, White Panelists, Nat'l L.J., Feb. 22, 1993, p. S8; Rita James Simon, The Jury and the Defense of Insanity 111 tbl. 45 (1967) (African Americans voted to acquit more than jurors of other races). But see Carol J. Mills & Wayne E. Bohannon, Character Structure and Jury Behavior: Conceptual and Applied Implication, 38 J. Personality & Soc. Psychol. 662, 666 (1980) (personality, not just demographics, shape juror decisions); Carol J. Mills & Wayne E. Bohannon, Juror Characteristics: To What Extent Are They Related to Jury Verdicts? 64 Judicature 22, 27 (1980) (reporting study finding African-American females more conviction prone).

[61]A recent racial breakdown of the number of judges in the four counties is:

	Total Judges	*Male*	*African American*	*Asian*	*Hispanic*
Bronx	29	22	4	0	4
DC	75	48	31	0	3
Los Angeles (criminal)	52	35	5	5	1
Maricopa	87	66	3	1	7

These are the number of judges sitting at the sites according to Garvey et al., supra note 39, n.8. Except for Los Angeles, the numbers in the table are not limited to judges sitting in criminal cases. In Maricopa County, 29 judges sat in the criminal division fiscal year 2000–2001. See note 18 supra.

[62]This excludes 412 jurors with missing race data.

Appendix: Site-Level Logistic Regressions of Conviction

	1 *Los Angeles*	2 *Maricopa*	3 *Bronx*	4 *DC*	5 *Los Angeles*	6 *Maricopa*	7 *Bronx*	8 *DC*
Judge	2.050*	–0.643	0.736	2.645***	0.716**	–0.163	0.556	1.948***
	(1.89)	(0.71)	(1.32)	(4.03)	(2.18)	(0.70)	(1.34)	(4.05)
Evidentiary strength—juries	3.390***	2.362***	0.566**	1.263***				
	(5.07)	(3.83)	(2.59)	(4.04)				
Evidentiary strength— judges	1.085***	2.386***	0.474***	0.312				
	(3.23)	(3.91)	(3.39)	(1.16)				
Male percent of jury	–4.717	0.132	0.249	–0.321	–0.052	2.240	0.181	–0.253
	(1.06)	(0.07)	(0.18)	(0.28)	(0.02)	(1.61)	(0.21)	(0.25)
African-American percent of jury	27.778***	–8.950	2.582**	–0.749	–1.009	1.868	1.975*	–1.868
	(3.68)	(0.80)	(2.08)	(0.47)	(0.36)	(0.38)	(1.70)	(1.21)
Hispanic percent of jury	4.894	2.375	0.892	–2.417	–4.840**	1.712	0.901	1.925
	(0.93)	(0.86)	(0.72)	(0.88)	(2.19)	(0.81)	(0.80)	(0.63)
Education level (1–5 scale)	2.798	–0.365	–1.020**	–0.574	–0.942	0.692	–1.006*	–1.137*
	(1.24)	(0.36)	(2.39)	(0.69)	(0.88)	(1.01)	(1.84)	(1.90)
Drug-possession charge dummy	2.081*	2.348***	–1.216		–0.328	0.722	–1.548	
	(1.68)	(2.88)	(1.31)		(0.41)	(0.91)	(1.59)	
Weapons charge dummy			–1.247**	–0.513			–1.482**	–1.561***
			(2.05)	(0.96)			(2.28)	(2.87)
Constant	–32.123***	–17.862***	–1.597	–3.657	6.590	–2.570	2.597	5.807*
	(2.94)	(3.69)	(0.87)	(0.96)	(1.60)	(0.98)	(1.48)	(1.92)
Observations	130	142	118	146	134	142	118	146

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

Notes: Absolute value of t statistics in parentheses, adjusted for multiple observations per case. NCSC data covering trials at four sites in 2000–2001. The number of cases is half the number of observations because each observation is duplicated to allow for a judge and jury outcome for each case.