

RUNNING HEAD: LIBERATION HYPOTHESIS

Strength of Evidence, Extraevidentiary Influence, and the Liberation Hypothesis:

Data from the Field

Dennis J. Devine, Jennifer Buddenbaum, Stephanie Houp

Indiana University Purdue University Indianapolis

Dennis P. Stolle

ThemeVision, LLC

Nathan Studebaker

Indiana University Purdue University Indianapolis

This article has been accepted for publication in *Law and Human Behavior*, the official journal of the American Psychology-Law Society/Division 41 of the American Psychological Association. The pre-publication original is available through the "Online First" service at www.springerlink.com.

Abstract

To examine relationships between strength of evidence (SOE) and extraevidentiary variables in the context of Kalven and Zeisel's (1966) liberation hypothesis, post-trial questionnaire data were collected from judges, attorneys, and jurors associated with 179 criminal jury trials. SOE ratings were strongly correlated with jury verdicts on the three most serious charges against the defendant, and several extraevidentiary variables (i.e., pretrial publicity, trial complexity, charge severity, and foreperson demographics) were moderately correlated with verdicts.

Extraevidentiary-verdict relationships remained significant when SOE was controlled, although extraevidentiary variables yielded only modest improvement in classification accuracy beyond SOE. In partial support of the liberation hypothesis, several case-related extraevidentiary variables were significantly related to jury verdicts only when the prosecution's evidence was rated as moderately strong.

Strength of Evidence, Extraevidentiary Influence, and the Liberation Hypothesis:

Data from the Field

In a landmark publication summarizing the results of perhaps the largest field study of jury decision-making ever conducted, Kalven and Zeisel (1966) offered an elegant and intuitive proposition about jury decision-making: When the evidence presented at trial clearly favors one side or the other, juries will usually base their decisions on the merits of the case and choose the corresponding verdict. However, when the evidence is ambiguous and does not definitively suggest an appropriate verdict, jurors will be “liberated” from its constraints and rely on other beliefs, assumptions, or sentiments about the participants in deciding. In essence, this so-called *liberation hypothesis* implies jury verdicts will be determined by the strength of the evidence in most trials but susceptible to non-evidentiary influences when the evidence is “close.”

Despite being well-known to jury scholars, there has been surprisingly little empirical research on the liberation hypothesis since it was offered more than 40 years ago. Indeed, only one study to our knowledge has examined it. Conversely, a large and growing body of experimental research on mock jurors has demonstrated that jury verdicts can be readily influenced by manipulating the quantity or nature of the evidence presented, as well as by other non-evidentiary variables that ideally should have no impact (see Devine, Clayton, Dunford, Seying, & Pryce, 2001; Greene, Chopra, Kovera, Penrod, Rose, Schuller, & Studebaker, 2002; Levett, Dannielson, Kovera, & Cutler, 2005 for reviews). The picture remains fuzzy, however, with regard to empirical relationships among strength of evidence, extraevidentiary variables, and verdicts outside the laboratory. This study was conducted to examine these relationships, as well as Kalven and Zeisel’s liberation hypothesis, using a sample of actual criminal jury trials.

Potential Determinants of Jury Verdicts

Strength of evidence. Even most critics of the jury system would probably concede that the quality and quantity of the evidence presented at trial plays a decisive role in most jury decisions. To the extent this is true, characteristics of the evidence should be strongly related to jury verdicts, and some research supports this proposition. Several studies have involved the coding of non-judgmental “surface” features of the evidence presented in actual trials, such as whether there was a confession by the defendant, the number of prosecution witnesses that testified at trial, whether the victim suffered serious physical injury, or whether a weapon used by the assailant was recovered. These non-judgmental indices of evidence presented have been found to be weakly-moderately associated with jury verdicts (e.g., Myers, 1979; Taylor & Hosch, 2004; Visher, 1987; Werner, Strube, Cole, & Kagehiro, 1985). However, “objective” characteristics of the evidence are limited in that they reflect only whether certain types of evidence were present and do not capture the subjective impact on jurors or how well the evidence presented fits together into a compelling whole.

Two studies, however, have featured the direct measurement of SOE in actual criminal trials and found it to be strongly related to jury verdicts (Devine, Olafson, Jarvis, Bott, Clayton, & Wolfe, 2004; Hannaford-Agor, Hans, Mott, & Munsterman, 2002). In both studies, global ratings of prosecution SOE based on absolute scales (e.g., “weak” to “very strong”) were obtained from judges, attorneys, and/or jurors via post-trial questionnaires. Devine et al. (2004) reported a correlation of .47 between conviction and prosecution SOE (as rated by lead attorneys), but this finding was based on only 14 jury trials. In the other study, a large-scale examination of hung juries in four different states conducted by the National Center for State Courts (NCSC), ratings of evidentiary strength were obtained from jurors, attorneys, and judges.

Based on one analysis of these data that was focused on explaining instances of jury nullification, Hannaford-Agor and Hans (2003) reported that “the weight and direction of the evidence” derived from juror responses was strongly related to the probability of acquittal verdicts. In another analysis of the same data set, Eisenberg et al. (2005) found that separate ratings of evidentiary strength obtained from the jurors and the judge were each strongly associated with the likelihood of conviction when included in a series of logistic regression models that also featured numerous other case characteristics. Thus, some research does indeed support the notion that characteristics of the evidence are related to actual jury verdicts.

Extrajudicial influence. Unfortunately, it is also clear that juries can be influenced by considerations other than the nature, quality, and quantity of evidence presented at trial. The impact of variables that should be irrelevant has been referred to variously as “extralegal” (e.g., Reskin & Visher, 1986; Visher, 1987), “legally impermissible” (e.g., Saks, 1997), and “extrajudicial” (e.g., Kerr, MacCoun, & Kramer, 1996; Levett et al., 2005). We will use the term *extrajudicial influence* to refer to any systematic effect on jury decision-making that does not stem directly from characteristics of the admissible evidence presented at trial.

Existing research suggests that several extrajudicial variables related to the nature of the case can influence jury verdicts. In particular, exposure to negative pretrial publicity (PTP) concerning the defendant or the case may prime jurors to be less critical of the evidence presented at trial and more likely to convict. This supposition is consistent with a meta-analysis of 44 empirical tests across 23 studies which revealed a positive association between exposure to PTP and judgments of the defendant’s culpability (Steblay, Besirevic, Fulero, & Jimenez-Lorente, 1999). The severity of the charge(s) against the defendant may also influence jury decision-making. Based on the assumption that jurors will seek to minimize the costs associated

with being wrong, the *severity-leniency hypothesis* (Kerr, 1978) predicts that the more severe the penalty prescribed for an offense, the lower the likelihood that the jury will convict. Consistent with this, several studies have found the likelihood of conviction to be inversely related to the length of the sentence associated with the charge (Kerr, 1978; Myers, 1979; Werner et al., 1985). In addition, trial complexity may also influence jury decisions (Heise, 2004; Heuer & Penrod, 1994; Mize, Hannaford-Agor, & Waters, 2007). *Legal complexity* refers to the quantity and density of the relevant law and the corresponding instructions given to the jury, whereas *evidentiary complexity* refers to the quantity and technicality of the testimony and exhibits presented at trial. High levels of either type of complexity might confuse or overwhelm some jurors, leading them to fall back on simple heuristics to identify a preferred verdict (e.g., a belief that most people arrested are guilty).

In addition to case-related characteristics, the personal characteristics of key trial participants (e.g., race, gender, socioeconomic status, physical attractiveness) may also influence jury verdicts (Devine et al., 2001; Levett et al., 2005). Many studies have been aimed at identifying general biases against defendants with particular characteristics, but some have also examined the possibility of more complex forms of bias based on the demographic similarity of jurors and other trial participants. In particular, an *ingroup-outgroup bias* (see Brewer, 2007 for a recent review) would make jurors more likely to acquit a defendant perceived as similar and convict a defendant who is not, whereas a *black sheep effect* (Marques, 1990) would involve a contingency determined by the strength of the evidence against the accused. Essentially, a black sheep effect would make jurors more likely to acquit a similar defendant when the evidence was weak, but more likely to convict when the evidence was strong. Acting punitively towards a

similar defendant facing strong evidence of guilt would essentially serve as a means for jurors to distance themselves from an ingroup member who had clearly violated group norms.

Although considerable research has been conducted on participant demographics in trial contexts, no clear pattern of results has emerged. In general, experimental research based primarily on non-interacting mock jurors has produced weak and/or inconsistent effects associated with participant race, gender, socioeconomic status, and physical attractiveness (Mazzella & Feingold, 1994; Sweeney & Haney, 1992; Mitchell, Haw, Pfeiffer, & Meissner, 2005). Conversely, a growing body of high-quality field research supports a connection between participant demographics and jury verdicts in actual trials, particularly those involving the death penalty (Baldus, Pulaski, & Woodworth, 1983; Baldus, Woodworth, Zuckerman, Weiner, & Broffitt, 1998; Bowers, Steiner, & Sandys, 2001). These studies are generally consistent with an ingroup-outgroup bias with regard to African-American defendants, but two mock jury studies have produced some support for the black sheep effect (i.e., Chadee, 1996; Kerr, Hymes, Anderson, & Weathers, 1995) and two field studies that examined bias associated with participant race did not produce clear support for either theory (i.e., Hannaford-Agor & Hans, 2003; Taylor & Hosch, 2004). In sum, it remains unclear when and how strongly participant demographics influence jury decisions.

Liberation Hypothesis

The liberation hypothesis represents a simple yet intuitive contingency theory about jury decision-making that can explain some of the mixed findings in the empirical literature. When the evidence presented at trial points clearly to a particular verdict, the jury will usually choose that verdict; when the evidence is more equivocal, the jury's decision will be more likely to reflect the influence of other normatively undesirable variables.

Although Kalven and Zeisel did not offer a cognitive explanation for the liberation hypothesis, theory and research on attitudes and persuasion is consistent with the idea that jurors may be most likely to fall victim to extraevidentiary influence when the prosecution's evidence is moderate in strength. In particular, dual-process theories of attitude formation and change posit the existence of two information-processing strategies, or "routes" (Eagly & Chaiken, 1993; Petty & Cacioppo, 1981; 1986). The *central route* involves systematic and effortful processing aimed at analyzing and evaluating messages based on the merits of their underlying arguments, whereas the *peripheral route* features a less intensive process wherein individuals rely primarily on heuristics and short-cuts to evaluate persuasive messages. In their elaboration likelihood model, Petty and Cacioppo (1986) argued that central processing is most likely to occur when decision-makers have ample time, are capable of understanding the arguments, and view the decision as important. These features probably characterize most jury trials. However, jurors attempting to sort out cursory, tangential, or conflicting evidence might be unable to choose a verdict using the logic and reasoning of central-route processing and become more susceptible to extraneous influences. Depending on the subsequent distribution of the individual verdict preferences and the dynamics of deliberation, extraevidentiary influences might be most decisive in cases like these with ambiguous evidence.

Despite its early articulation, only one study to our knowledge has pointedly investigated the liberation hypothesis. Based on direct courtroom observation and post-trial interviews with jurors in 38 sexual assault cases, Visher and Reskin found that several objective characteristics of the evidence accounted for variation in individual jurors' pre-deliberation verdict preferences (Reskin & Visher, 1986; Visher, 1987). However, "extralegal" factors associated with participant demographics (i.e., defendant, juror, and victim) were also related to pre-deliberation judgments

and improved prediction models when entered after the evidence characteristics. Further, when separate analyses were done for jurors exposed to cases with strong as opposed to weak evidence, extralegal factors tended to have larger regression coefficients in the analysis of weak cases. Although the analyses focused solely on individual pre-deliberation judgments in a relatively small sample of sexual assault trials and did not distinguish cases with moderate SOE levels, this research does provide some preliminary support for the liberation hypothesis.

Present Study

Kalven and Zeisel's (1966) liberation hypothesis suggests extraevidentiary variables will be most influential when the prosecution's evidence is only moderately strong and does not clearly favor either side. Experimental research based mostly on mock jurors shows that SOE and extraevidentiary variables can influence mock jury verdicts but, with the exception of the recent NCSC study of hung juries, no study has involved the direct measurement of SOE *and* a broad set of extraevidentiary variables in real juries. Further, there has been no direct jury-level test of the liberation hypothesis using a large number of actual verdicts or a diverse set of charges. To address these issues, we conducted a field study of criminal juries to examine the extent to which: (1) SOE and extraevidentiary variables are related to jury verdicts in actual criminal trials, (2) extraevidentiary variables can explain jury verdicts after taking SOE into account, and (3) extraevidentiary variables are more strongly related to jury verdicts when SOE is moderate as opposed to extreme (i.e., clearly favoring one side). We anticipated that both SOE and extraevidentiary variables would be significantly associated with jury verdicts and that extraevidentiary variables would improve the classification of jury verdicts beyond SOE. Further, in accord with the liberation hypothesis, we expected extraevidentiary-verdict relationships to be strongest when SOE was moderate.

Method

Participants

After securing the permission of the court's executive committee, all 18 trial court judges of the Superior Court of Marion County (Indianapolis) were invited to participate in the study at a meeting of the criminal division judges attended by the first author. In addition, to increase the diversity of the sample of cases, several judges in Allen County (Fort Wayne) were also invited to participate. In total, 10 criminal court judges participated in the study, nine of whom were associated with Marion Superior Court.

We eventually received 878 juror questionnaires, 149 judge questionnaires, 95 lead prosecuting attorney questionnaires, and 76 lead defense attorney questionnaires from 195 criminal trials. This corresponds to response rates of 77% for participating judges, 49% for lead prosecutors, and 39% for lead defense attorneys. At least one legal professional (i.e., judge or attorney) responded for 169 of the trials, two or more responded for 108 trials, and all three legal professionals responded for 43 trials ($M = 1.64$, $SD = .97$). In total, we obtained data from 28 six-person criminal jury trials and 166 twelve-person criminal jury trials. The mean number of responding jurors was 2.79 ($SD = 1.50$) for the 6-person juries and 4.82 ($SD = 1.89$) for the 12-person juries, yielding response rates of 47% for 6-person juries and 40% for 12-person juries.

Measures

Data were collected via four paper-and-pencil questionnaires, each customized for a particular type of respondent (i.e., judge, lead prosecuting attorney, lead defense attorney, juror). To minimize the burden on respondents and thus maximize response rate, we limited each questionnaire to the two sides of a single 8.5" x 11" piece of paper. To allow both trials and respondents to remain anonymous, no identifying information about either was requested.

Jury verdicts for each charge were recorded by jurors using three nominal categories (i.e., Not Guilty, Hung, Guilty). A dichotomous version for primary use in the analyses was created by combining acquittals and hung juries (0 = Not Guilty/Hung, 1 = Guilty), and an alternative version of this was generated by deleting those cases with hung verdicts (0 = Not Guilty; 1 = Guilty). To generate an index of within-jury agreement on the verdict, the number of responding jurors who agreed on the jury's final verdict for each charge was divided by the total number of responding jurors. The mean of the resulting within-jury agreement values was 92% for Charge 1, 85% for Charge 2, and 81% for Charge 3, indicating very good agreement on the focal trial outcomes. In other words, for a typical jury that yielded five respondents, four or more usually reported the same verdict for any given charge. These estimates are also somewhat conservative in that each juror did not always report a verdict for each of the three most severe charges and our calculation method does not distinguish between disagreement and missing data.

Strength of evidence (SOE) ratings were obtained by asking the two lead attorneys and the judge to rate the strength of the *prosecution's* case on each of the three most serious charges using a 6-point response scale (1= Weak, 6 = Very Strong). An overall rating of SOE was also obtained using the same scale. Pairwise correlations among the three rating sources were consistently strong and positive (generally between .4 and .6), and internal consistency analyses in which the three raters were treated as separate "items" revealed good reliability associated with the mean SOE ratings ($\alpha = .80, .77, \text{ and } .79$ for Charges 1, 2, and 3, respectively). In light of this convergence, jury-level prosecution SOE scores were calculated for each charge by taking the mean of the available ratings.

Prosecution evidence presented was indexed separately from SOE in a checklist fashion by asking prosecuting attorneys to indicate which of the following types of evidence were

presented by the State during the trial: (1) eyewitness testimony, (2) arresting officer testimony, (3) expert testimony, (4) ballistic tests, (5) DNA tests, (6) other physical evidence, (7) recovered personal items, (8) a recovered weapon, (9) crime scene photos/videotape, and (10) other notable evidence. For each jury trial, a score was generated by summing the number of checks, yielding possible scores ranging from 0 to 10.

Charge severity was coded from information about the three most serious charges against the defendant provided by all respondents using a fill-in-the-blank format. Codes were based on an 8-point classification scale corresponding to the severity of the penalty for conviction under Indiana statutory law (1= Class C Misdemeanor; 8 = Murder). In many trials, there was only one charge or there were multiple charges that fell in a clear descending order with regard to severity. Occasionally, however, defendants faced multiple charges at the same severity level (e.g., Class B Felony), in which case the charge that produced the highest level of agreement on the jury's verdict was designated the more severe charge. If the same level of juror agreement existed for multiple charges at the same severity level, we followed the ordering provided by a plurality of the respondents. Overall, 25% of the trials involved charges with the same severity levels for Charges 1 and 2 (e.g., Felony A), 12% involved charges with the same severity for Charges 2 and 3, and 6% involved charges with the same level for all three of the most severe charges.

Pretrial publicity was indexed by asking judges to respond to the following two items: (1) Were you aware of any significant pretrial publicity related to the case in the media?, and (2) Was any juror discharged during voir dire for having been exposed to media coverage of the case? Both items featured a dichotomous response scale (0 = No, 1 = Yes).

Trial complexity was measured by asking judges and the two lead attorneys to rate two items, *legal complexity* and *evidentiary complexity*, using a 4-point scale (1 = Simple, 4 = Very Complex). We initially coded the two dimensions of complexity separately by calculating the mean of the available responses for each dimension of complexity for each respective trial. However, internal consistency analyses treating responses from the three rating sources as items produced alpha values that were somewhat low ($\alpha = .60$ for legal complexity; $\alpha = .55$ for evidentiary complexity). Further, the two complexity dimensions were strongly and positively correlated with each other and displayed a similar pattern of relationships with other study variables. Accordingly, we opted to combine ratings for the two dimensions of complexity into a more general index of *trial complexity*. Trial complexity scores for a particular trial were calculated by taking the mean of the legal and evidentiary complexity ratings available, yielding an adequate coefficient alpha internal consistency ($\alpha = .70$).

Attorney competence was measured by asking jurors to rate the competence of the State's attorneys and the defense's attorneys (separately) using one 5-point Likert-scale item. Attorneys and judges were not asked to rate themselves or each other because of the sensitive nature of the question and the obvious potential for leniency bias, as well as concerns on the part of judges that such data might influence appellate proceedings if they should ever become public. The mean within-jury rating was thus used as a rough index of the skill of counsel in each of the jury trials. Treating "jury" as the independent variable and individual juror ratings as the dependent variable, intra-class correlation coefficient (ICC) values were calculated to assess the reliability of the mean within-jury rating (Bliese, 2000). These values were .69 for both the State and the defense.

Defendant characteristics were provided by jurors, and in some cases by judges and attorneys as well. Jurors were asked to indicate the defendant's *gender* (0 = Female, 1 = Male), *race* (1 = Black, 2 = Hispanic, 3 = White, 4 = Asian-American, 5 = Other), and *testimony* (0 = No, 1 = Yes), with defendant race subsequently recoded into a dichotomous variable (0 = Non-white; 1 = White) for the primary analyses. The modal value provided by responding jurors was used for the trial, and these values were cross-checked with values provided by the judge and defense attorney where possible for gender and race (no discrepancies were found). The mean percentage of responding jurors who agreed on the modal value was 97% for both defendant gender and defendant race. Defendant *likeableness* and *attractiveness* scores were generated for each trial by taking the mean of one 5-point Likert-scale item completed by jurors. This yielded ICC values of .57 for defendant likeability and .50 for defendant attractiveness.

Jury demographic composition was coded from information provided by responding jurors, who were asked to indicate the number of men and women on their jury in each of the following groups: Hispanic, Black, White, Asian-American, Middle-Eastern, and Other. Jury-level values for each demographic category were identified by taking the modal number of jurors reported in that category by responding jurors. To assess agreement on demographic composition, we counted the number of responding jurors who agreed on *every* element in the modal demographic profile and divided that value by the total number of responding jurors. For example, if the most common demographic profile for a jury was 4 White males, 5 White females, and 3 Black females, we counted the number of responding jurors who indicated this exact distribution. Thus, the reported agreement values are somewhat conservative in that jurors who dissented on *any* element of the profile were not counted as agreeing. In general, there was good agreement on jury demographic composition, with an average of 72% of responding jurors

agreeing on the modal profile. Several additional indices reflecting jury-defendant similarity were also derived from these basic data. *Jury-defendant racial similarity* and *jury-defendant gender similarity* were created by counting the number of jurors who matched the defendant's race *or* gender (e.g., White jurors, female jurors); *jury-defendant composite similarity* was indexed by counting the number of jurors who matched the defendant's race *and* gender (e.g., Black male jurors).

Foreperson race and **foreperson gender** were measured by having jurors circle the appropriate category describing their foreperson from the same set of demographic choices used for defendants. Mean within-jury agreement was excellent for both variables (i.e., 97% for foreperson gender and 95% for foreperson race). Responding jurors were also asked to indicate if they had been the foreperson and, when this person responded, these “true scores” were cross-checked with modal values and no discrepancies were found.

Procedure

Packets of judge, attorney, and juror questionnaires were initially assembled and delivered (or mailed) to participating courts, and then additional packets were supplied as needed. Each questionnaire was pre-coded with a three-digit “Trial ID” number used to link the responses associated with a given trial. The protocol for survey administration was as follows: After each trial and while the jury was deliberating, participating judges were asked to read a brief description of the study to the attorneys and invite the lead attorney for each side to participate in the study. Judges and attorneys were asked to complete their questionnaire as soon as possible, ideally before jury verdicts had been announced; approximately 40% of the legal professionals were able to do so (i.e., 46% of judges, 33% of prosecuting attorneys, and 42% of defense attorneys). After the jury returned and announced its verdict(s), judges were asked to

read the study description to the assembled jurors and invite them to participate as well. Judges were instructed to announce the study and distribute study questionnaires at each jury trial they presided over during the 21 months of active data collection.

After active data collection ceased, a member of the research team examined all the returned questionnaires for a given trial and filled out a master jury-level coding sheet for all categorical variables after determining the ordering of the charges with regard to severity. All primary analyses were conducted at the *jury* level, with jury-level scores generally set equal to the mean (for continuous variables) or the mode (for categorical variables) of the available responses. When there was no mode present or multiple modes for a categorical variable, it was treated as “missing” (although this was rare due to the high levels of within-jury agreement).

Results

Descriptive Data

Overall, at least one response was received from 195 criminal jury trials and usable data were available for 179 juries (21 with six members, 158 with twelve members). The data from a jury trial were considered “usable” if the trial featured a complete jury deliberation and we received responses from two or more jurors who agreed on at least one verdict. Sixteen cases were lost from the data set due to: (1) their being from trials that ended before or during deliberation (i.e., mistrial, directed verdict, late plea bargain), (2) receiving questionnaires from fewer than two jurors, or (3) failure of responding jurors to agree on at least one verdict.

Table 1 presents means, standard deviations, and predictor-verdict correlations for focal study variables in the 179 usable criminal jury trials. The conviction rate was 52% for the most serious charge, 58% for the second most serious charge, and 67% for the third most serious charge. Across the three charges, defendants were found guilty on at least one charge in 62% of

the trials. Defendants were most often Black (71%) and male (94%), and testified in 42% of the trials. Judges reported a significant amount of pretrial publicity in the media prior to 19 of the 138 trials (14%) for which a judge questionnaire was available, and in 7 of the 138 trials (5%) they indicated at least one juror was dismissed during voir dire for exposure to such publicity. Almost all juries had a White majority (83%) or were composed entirely of White jurors (16%). Half of the juries had a female majority, 29% had a male majority, and 21% were evenly split between men and women. Forepersons were usually White (87%) and male (71%). Mean ratings of the prosecution's strength of evidence fell between moderate and strong ($M_1 = 4.23$, $M_2 = 4.39$, $M_3 = 4.34$) and the trials on average were rated as modestly complex ($M = 1.93$ for legal complexity; $M = 1.78$ for evidentiary complexity), although research has shown that legal professionals tend to rate trials as less complex than jurors (Heise, 2004; Mize et al., 2007).

Correlations with Jury Verdicts

A primary goal of this study was to examine the degree to which SOE and variables representing potential extraevidentiary influence are associated with jury verdicts. In keeping with convention, hung juries were combined with acquittals to form a dichotomous outcome variable (i.e., 0 = hung/acquit, 1 = convict). Unless otherwise indicated, correlations discussed in this section were statistically significant at $p \leq .01$, and the subscripts "1," "2," and "3" denote correlations involving the three most serious (i.e., severe) charges, with Charge 1 denoting the most serious charge and Charge 3 the least serious. Many trials did not involve multiple charges, resulting in lower sample sizes for the less severe charges and a large influence on corresponding statistical significance levels.

In terms of variables related directly to the evidence and the attorneys' handling of it, there was a substantial positive correlation between conviction and the strength of evidence

rating for all three of the most severe charges ($r_1 = .48, r_2 = .60, r_3 = .43$), with a guilty verdict more likely when the prosecution's evidence was rated as stronger by the attorneys and judge. The diversity of evidence presented by the prosecution was also positively related to jury verdicts for all three charges ($r_1 = .39; r_2 = .44; r_3 = .40, p = .03$) but was not strongly correlated with SOE ratings for any of the three charges ($r_1 = .28; r_2 = .27, p = .04; r_3 = -.06, ns$). In other words, indices of evidence presented did not appear to be particularly good proxy variables for direct measures of strength of evidence. Further, prosecuting attorney competence as rated by the jurors was more strongly associated with jury verdicts ($r_1 = .35, r_3 = .38$) than the corresponding rating for defense attorney competence ($r_1 = -.16, p = .04$). Thus, variables that logically should be related to jury verdicts were, in fact, fairly good predictors.

Consistent with previous research on mock jurors, there were also signs of potential influence associated with several extraevidentiary characteristics pertaining to the nature of the case. Specifically, charge severity was positively related to conviction on the most serious charge ($r_1 = .32$) as was pretrial publicity ($r_1 = .26$), whereas trial complexity was positively associated with guilty verdicts on all three charges ($r_1 = .33, r_2 = .26, r_3 = .44$). In essence, guilty verdicts were more likely when: (a) the defendant was faced with a more serious charge, (b) substantial pretrial publicity was reported by the judge, and/or (c) the trial was rated as more complex by the legal professionals in terms of the evidence and relevant law. In contrast to these case-related characteristics, most demographic variables associated with the defendant and jurors were not significantly associated with jury verdicts. Only one demographic variable associated with the jury's overall demographic composition was significantly related to jury verdicts -- the percentage of Black female jurors for Charge 2 ($r_2 = -.27$), with conviction less likely in juries with more Black female jurors. However, the foreperson's race ($r_1 = .20$) and gender ($r_1 = .17, p$

= .02) were related to jury verdicts for Charge 1, with convictions more likely when forepersons were White (27% higher conviction rate) or male (15% higher conviction rate).

As a check, we also calculated the correlations in Table 1 using: (a) only juries that reached a verdict (i.e., excluding hung juries), and (b) an additional dichotomous verdict measure reflecting conviction on *any* (i.e., one or more) of the three most severe charges. These analyses produced a pattern of results very similar to that evident in Table 1, with small (i.e., $< \pm .05$) and unsystematic differences in corresponding correlations as a function of including/excluding hung juries or verdict measure.

Correlations with Jury Verdicts Controlling for SOE

The right side of Table 1 displays partial correlations (pr) between jury verdicts and predictor variables controlling for the charge-specific SOE presented by the prosecution. In general, case characteristics significantly related to verdicts in a zero-order fashion remained so at similar levels when SOE was controlled. Of particular note, after removing the variance attributable to SOE, higher levels of charge severity ($pr_1 = .30$; $pr_2 = .22$, $p = .04$), pretrial publicity ($pr_1 = .22$), and trial complexity ($pr_1 = .34$; $pr_2 = .25$, $p = .04$; $pr_3 = .49$) were all still positively and significantly associated with conviction. Further, foreperson race ($pr_1 = .20$, $p = .02$) and foreperson gender ($pr_1 = .16$, $p = .04$) also remained significantly related to Charge 1 verdicts, as did the percentage of Black female jurors with Charge 2 verdicts ($pr_2 = -.23$, $p = .03$). Thus, a number of extraevidentiary variables associated with the case and key participants remained related to jury verdicts even after the State's SOE was taken into account. In addition, the index of the prosecution's evidence presented was also positively related to conviction after controlling for SOE ($pr_1 = .30$, $pr_2 = .35$, $pr_3 = .46$, $p = .02$), indicating that trials in which the

prosecution presented more *types* of evidence tended to yield more guilty verdicts independent of the perceived *strength* of that evidence.

Multivariate Classification of Jury Verdicts

To address the question of whether extraevidentiary variables can improve the explanation of jury verdicts beyond the evidence, we conducted a hierarchical multivariate logistic regression analysis involving SOE, a set of extraevidentiary variables, and Charge 1 verdicts. For this analysis, we included all extraevidentiary variables significantly related to Charge 1 jury verdicts in a bivariate fashion. SOE was entered at Step 1 and all five extraevidentiary variables were entered as a set at Step 2 (i.e., charge severity, pretrial publicity, trial complexity, foreperson race, foreperson gender). We limited this analysis to Charge 1 verdicts due to the smaller sample sizes available for Charges 2 and 3.

Table 2 displays the results of the hierarchical logistic regression analysis. The base rate for correct classification of jury verdicts was 52% (i.e., the percentage of juries that convicted). When entered at Step 1, SOE produced a statistically significant improvement in the model, $\chi^2(1) = 34.23$, $Wald = 24.82$, $p < .01$; $OR = 2.85$; Nagelkerke $R^2 = .30$, along with a 25% increase in classification accuracy to 77.2%. The positive regression coefficients for all the predictors indicate that higher values of each predictor were associated with guilty verdicts, with the odds ratio (OR) values reflecting the increase in the odds of conviction for each one-unit increment on the predictor. Thus, for every one-unit increase in the mean prosecution SOE rating, the odds of a guilty verdict roughly tripled. At Step 2, the five extraevidentiary variables collectively produced a significant improvement in the prediction model, $\Delta\chi^2(5) = 33.38$, $p < .01$; Nagelkerke $R^2 = .52$, and a small increase in classification accuracy to 80.1%. SOE remained significant at Step 2 ($Wald = 21.04$, $p < .01$; $OR = 3.10$), as did foreperson race ($Wald = 5.45$, p

= .02; $OR = 9.04$), and trial complexity ($Wald = 6.18, p = .01; OR = 2.64$). Foreperson gender just missed the traditional .05 level of significance ($p = .06$). In essence then, after accounting for prosecution SOE, the odds of conviction were substantially higher for each additional point on the trial complexity scale, and when the jury had a White (as opposed to non-White) foreperson.

As a check on the stability of these results, we also conducted two alternative versions of the main logistic regression analysis in which: (a) hung juries were excluded, and (b) a dichotomous verdict measure reflecting conviction on *any* of the three most serious charges was employed as the outcome in conjunction with an overall SOE rating for the State's evidence. In both alternative analyses, trial complexity was the only extraevidentiary variable with a significant regression coefficient at Step 2, although foreperson race was marginally significant when hung juries were excluded. In addition, we also ran a set of hierarchical logistic regression analyses to examine potential interactions between each extraevidentiary variable and SOE. In this set of regressions, dichotomous jury verdict on Charge 1 served as the dependent variable, SOE was entered at Step 1, a single extraevidentiary variable was entered at Step 2, and the interaction between SOE and the extraevidentiary variable was entered at Step 3. None of the interactions were found to be significant at Step 3. Also, to rule out the possibility that the foreperson race effect was simply the manifestation of "classic" prejudice associated with many Black defendants and White-majority juries, we ran a logistic regression in which Charge 1 verdicts were regressed on SOE (Step 1), percentage of white jurors and defendant race (Step 2), and foreperson race (Step 3). The regression coefficient for foreperson race was significant at Step 3 ($p < .01$) and accompanied by a large change in the odds ratio (9.20), whereas defendant race and the percentage of white jurors both failed to produce large or significant effects at Steps 2 or 3.

Liberation Hypothesis

Finally, to examine Kalven and Zeisel's liberation hypothesis, we also calculated correlations between extraevidentiary variables and Charge 1 verdicts after sorting all trials into one of three evidentiary categories: *pro-acquittal* (SOE < 4.00; n = 45), *ambiguous* (SOE between 4.00 and 4.99; n = 52), and *pro-conviction* (SOE rating > 4.99; n = 59). Results are shown in Table 3.

The general pattern of results provides partial support for the liberation hypothesis. Consistent with the hypothesis, several correlations between potential sources of extraevidentiary influence were strong and significant for cases classified as ambiguous that featured moderate prosecution SOE. These extraevidentiary variables were charge severity ($r = .32, p = .02$), pretrial publicity ($r = .39$), trial complexity ($r = .54$), and foreperson race ($r = .35; p = .02$). On the other hand, defendant gender ($r = -.47$ for pro-acquittal cases) and foreperson gender ($r = .30, p = .05$ for pro-acquittal; $r = .39$ for pro-conviction) were significantly related to jury verdicts when the evidence was more definitive, but not when SOE was in the moderate range. In addition, there was little support for an ingroup-outgroup (similarity-leniency) bias in these data, with weak and *positive* (although non-significant) correlations between jury verdict and the number of jurors matching the defendant's race and/or gender across all three evidentiary categories. Consistent with the black-sheep effect, however, positive correlations around .2 were observed for all three jury-defendant similarity variables when the evidence clearly favored conviction, although none of these correlations attained statistical significance at the .05 level. Overall, these data provide some support for the liberation hypothesis in suggesting that extraevidentiary influence related to the case itself (i.e., charge, pretrial publicity, trial complexity) may indeed be greater when the prosecution's evidence on the most serious charge

is only moderate in strength, but this pattern does not appear to hold for demographic variables related to the defendant and foreperson.

Discussion

Summary

Five major findings emerge from the results of this study: (1) charge-specific ratings of prosecution SOE were strongly related to corresponding jury verdicts, (2) a number of extraevidentiary variables associated with the trial context and participant demographic characteristics were also moderately associated with jury verdicts, (3) after controlling for charge-specific SOE, significant relationships remained between several extraevidentiary variables and jury verdicts, (4) extraevidentiary variables only modestly improved classification of jury verdicts over SOE on the most serious charge, and (5) several trial-related sources of extraevidentiary influence followed a pattern consistent with Kalven and Zeisel's liberation hypothesis in being significantly correlated with verdicts on the most serious charge when the evidence was ambiguous, but not when it clearly favored one of the two sides.

Strength of Evidence

In light of persistent concerns and outright skepticism regarding jury performance, it was notable that charge-specific SOE correlated .48 to .62 with jury verdicts and improved the classification of jury verdicts by 22% over the base rate of conviction on the most serious charge. This study thus replicates earlier research using overall estimates of SOE (i.e., Devine et al., 2004; Hannaford-Agor & Hans, 2003) and extends that work by distinguishing specific charges and verdicts, as well as the strength of the evidence relevant to each. Many criminal jury trials involve multiple charges against the defendant, yet this is the first study to our knowledge to obtain separate SOE measures for each charge rather than using one global estimate. In reality,

SOE is likely to vary across a set of charges against a defendant, being perhaps quite strong for some charges and rather weak for others. In this study, charge-specific ratings of SOE were strongly related to jury verdicts across all three charges, and correlations involving charge-specific SOE were somewhat larger than corresponding correlations involving overall SOE ratings (with the exception of Charge 3). Future studies seeking to incorporate SOE could benefit from the enhanced precision associated with distinguishing SOE and verdicts by charge.

Given the prior use of objective measures of evidence presented in a number of archival data sets, it was also interesting to observe that correlations between the prosecution's evidence-presented index and the jury verdicts were not much lower than the SOE-verdict correlations. Further, the prosecution's evidence presented index remained significantly related to jury verdicts when SOE was controlled. In other words, independent of the perceived strength of the prosecution's evidence, having a greater variety of it appeared to provide some additional advantage to the State. It may be that jurors feel more comfortable convicting if all of their evidentiary eggs are not in the same basket. Future field studies focusing on the role of the evidence should distinguish and obtain non-judgmental indices of evidence presented as well as ratings of perceived evidence strength from knowledgeable observers.

Extrajudicial Influence

Three extrajudicial variables associated with aspects of the case were also moderately related to jury verdicts – pretrial publicity, charge severity, and trial complexity. The finding that pretrial publicity had a moderate positive relationship with conviction in this study involving actual juries is consistent with the Steblay et al. (1999) meta-analysis based primarily on experimental studies involving mock jurors acting as individuals. Indeed, that meta-analysis produced an overall effect-size measure for PTP (mean $r = .15$) at the low end of the effect-size

range we obtained in this study with real juries (.15-.30, depending on the charge and whether SOE was controlled). As a result, this study provides some corroboration from the field regarding the negative impact of pretrial publicity on a defendant's chances in front of a jury. It is important to note, though, that we did not measure *exposure* to PTP in this study – simply judges' awareness of its *existence*. It would thus be beneficial to replicate this finding with direct measurement of jurors' PTP exposure in future research.

In contrast to PTP, the findings regarding charge severity run counter to the literature and the “severity-leniency” hypothesis identified by Kerr (1978). This hypothesis states that juries deciding on more serious charges will tend to be more lenient (i.e., acquit at a higher rate) in light of the harsher consequences for the defendant if they mistakenly convict. Our data present a somewhat contradictory picture regarding this hypothesis. On one hand, looking *across defendants*, conviction tended to be more likely on all three charges when they were more severe (i.e., a higher felony), even when SOE was controlled. However, looking across charges *within defendants*, the likelihood of conviction increased systematically from Charge 1 to Charge 3. One explanation for this seeming paradox is that, other things being equal, more severe charges do pre-dispose jurors to believe the defendant probably did something that warrants punishment but, when faced with multiple charges against a defendant, juries still show some leniency by acquitting on the most serious charge and convicting on one or more of the less serious ones.

These results are also consistent with the notion that jurors faced with highly technical evidence or lengthy, convoluted instructions might somehow translate a sense of overload into a belief that the evidence points to conviction. Trial complexity was correlated at .33 with conviction on the most serious charge, remained at the same level when SOE was controlled, and rose to a whopping .54 when the prosecution's evidence was rated moderate to strong. Of note,

the separate correlations for legal complexity and evidentiary complexity (not presented) were remarkably similar to each other, suggesting that the two primary dimensions of the complexity construct operate in a similar fashion and the focus of concern cannot be narrowed to one or the other. Overall, these data are consistent with the notion that high levels of trial complexity may induce jurors to fall back on assumptions or heuristics that favor the prosecution (e.g., those who get arrested are probably guilty).

In contrast to the extraevidentiary variables associated with characteristics of the case itself, the picture regarding bias related to participant demographics is less clear. Two guiding theoretical perspectives have emerged in recent years (i.e., ingroup-outgroup bias, and the black sheep effect), but the results of this study do not provide much support for either perspective. Rather, these results are most consistent with the normative ideal in that juror and defendant demographic variables were not strongly related to jury verdicts – with a few exceptions. In particular, juries with White or male forepersons were more likely to convict on the most serious charge. Foreperson race remained significant and produced a large change in the odds ratio in the logistic regression that included both SOE and the other significant extraevidentiary variables, providing some support for the robustness of the effect. Although there has been a fair amount of research on foreperson characteristics (see Devine et al., 2001 for a summary), we know of no other field study that found foreperson race or gender to be related to jury verdicts. One potential explanation for a general tendency on the part of juries with White and male forepersons to convict is that such individuals are more likely to hold privileged places in the community, and those who emerge as leaders in a jury may feel called to serve as gate-keepers who protect society. In this role, they might react strongly to *any* person seen as posing a threat to the social order – regardless of their skin color or gender.

Finally, it was notable that the five extraevidentiary variables significantly related to Charge 1 verdicts collectively improved the classification accuracy by only 3% after SOE was taken into account. Given the variety of extraevidentiary variables represented and the strength of their bivariate relationships, this increase seems rather modest. Thus, these results do suggest that extraevidentiary variables can improve the explanation of jury verdicts beyond SOE, but not substantially.

Liberation Hypothesis

This is the first study to our knowledge to examine the liberation hypothesis using direct measures of SOE and actual jury verdicts. Reskin and Visher (1986) examined the hypothesis by focusing on juror pre-deliberation verdict preferences at the individual-level in 38 sexual assault cases and using two case evidence categories (i.e., weak and strong). We examined extraevidentiary-verdict relationships in real juries using three categories based on SOE levels for the most serious charge (pro-acquittal, ambiguous, pro-conviction). Although caution is warranted due to the moderate number of cases in each evidentiary category, the overall pattern was interesting: Case-related extraevidentiary variables (i.e., charge severity, pretrial publicity, trial complexity) followed the pattern suggested by Kalven and Zeisel in being significantly correlated with verdicts when the prosecution's evidence was only moderately strong, whereas participant demographic variables (primarily foreperson race and gender) did not operate as expected. One potential explanation for this discrepancy is that bias associated with foreperson demographic characteristics is simply very robust and manifests itself regardless of how strong the evidence is. Alternatively, the anomalous results for participant demographics may be due simply to sampling error associated with the small numbers of certain demographic types (e.g., non-White forepersons, female defendants) in some of the three categories. For example, the

observed relationship between defendant gender and jury verdicts for pro-acquittal cases ($r = -.47$) is attributable in large part to the fact that all three female defendants in that category were convicted. Overall, these results provide some support for the liberation hypothesis insofar as case-related extraevidentiary variables are concerned (i.e., pretrial publicity, trial complexity, and charge severity), but the mixed pattern of results suggests the need for further study of Kalven and Zeisel's liberation hypothesis to determine if demographic bias operates differently than case-related extraevidentiary influence.

Study Limitations

In addition to the qualifications already noted, several general considerations should be kept in mind when interpreting the results of this study. First, it is possible that observed relationships may have been influenced to some degree by "retrospective rationality" on the part of the respondents. However, most study variables were factual in nature, based on the reports of multiple respondents, and with a few exceptions exhibited very high levels of within-jury agreement. Second, given that response rates for attorneys and jurors ranged from 30-50%, the responses we received for any given trial may have differed in some systematic fashion from the set of potential responses. Third, judges were instructed to announce the study and distribute the questionnaires at every trial they presided over during data collection, but it is unknown how well judges followed that instruction. Thus, the sample of trials we obtained might have differed in some ways from the population of jury trials that occurred in the collective courtrooms of participating judges. If anything, though, participating judges seemed over-eager to generate data for the study, as we learned when we received completed questionnaires back from several non-eligible trials (e.g., mistrials, directed verdicts, or last-minute plea bargains during the trial). Fourth, inferences related to charge severity are qualified by the existence of some ties in

severity level across charges. Finally, most of the data for this study came from state courts in one jurisdiction (Marion County, Indiana), and so it is possible that some of these findings might not generalize to other jurisdictions. This possibility is tempered somewhat, however, by the participation of multiple judges and the wide variety of charges featured in this study.

Conclusion

There has been considerable debate about the jury system over the last half-century. A fundamental issue underlying this controversy is the empirical question of how much juries are influenced by extraneous considerations unrelated to the evidence. This study examined data from a large sample of actual criminal jury trials and featured the direct measurement of SOE along with a wide variety of potential sources of extraevidentiary influence. The results suggest both good news and bad news. In terms of the good, the strength of the prosecution's evidence was a strong predictor of jury verdicts for all three of the most serious charges against the defendant, and extraevidentiary variables as a set only marginally improved the classification of jury verdicts. In terms of the bad news, several extraevidentiary variables remained moderately related to jury verdicts even after taking SOE into account. Further, as predicted by the liberation hypothesis, extraevidentiary-verdict relationships involving the case itself (i.e., pretrial publicity, charge severity, trial complexity) were strongest when SOE was moderate – but relationships involving foreperson race and gender did not follow this pattern. Thus, with regard to Kalven and Zeisel's intriguing liberation hypothesis, additional research featuring a large sample of jury trials from multiple jurisdictions would be helpful to determine if ambiguous evidence is indeed truly liberating.

Author Note

Dennis J. Devine, Jennifer Buddenbaum, and Stephanie Houp, Department of Psychology, Indiana University Purdue University Indianapolis; Dennis P. Stolle, ThemeVision LLC, Indianapolis, Indiana; Nathan Studebaker, Department of Psychology, Indiana University Purdue University Indianapolis.

Jennifer Buddenbaum is now at the Indiana University School of Medicine, Children's Health Services Research. Nathan Studebaker is now at *pan – a TALX company*.

A report of project findings focused primarily on deliberation has been published in the *Journal of Empirical Legal Studies* (2007), Volume 4, 273-303. We thank the judges and staff of the Superior Court of Marion County, Indiana, as well as the judges and staff of the Superior Court of Allen County, Indiana, for participating in this study. We also thank Julie Devine for her editorial assistance in preparing the manuscript.

Correspondence concerning this article should be addressed to Dennis J. Devine, LD 124, Department of Psychology, IUPUI, 402 N. Blackford St., Indianapolis, Indiana, 46202. E-mail: ddevine@iupui.edu.

References

- Baldus, D.C., Pulaski, C., & Woodworth, G. (1983). Comparative review of death sentences: An empirical study of the Georgia experience. *Journal of Criminal Law and Criminology*, 74, 661-753.
- Baldus, D.C., Woodworth, G., Zuckerman, D., Weiner, N.A., & Brofitt, B. (1998). Racial discrimination and the death penalty in the post-Furman era: An empirical and legal overview, with recent findings from Philadelphia. *Cornell Law Review*, 83, 1638-1661.
- Bliese, P. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K.J. Klein and S.W.J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349-381). San Francisco: Jossey-Bass.
- Bowers, W.J., Steiner, B.D., & Sandys, M.R. (2001). Death sentencing in Black and White: An empirical analysis of the role of jurors' race and jury racial composition. *University of Pennsylvania Journal of Constitutional Law*, 3, 171-274.
- Brewer, M.B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist*, 62, 728-738.
- Chadee, D. (1996). Race, trial evidence and jury decision making. *Caribbean Journal of Criminology and Social Psychology*, 1, 49-86.
- Devine, D.J., Clayton, L.D., Dunford, B.D., Searing, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law*, 7, 622-727.

- Devine, D.J., Olafson, K.M., Jarivs, L.L., Bott, J.P., Clayton, L.D., & Wolfe, J.M.T. (2004). Explaining jury verdicts: Is leniency bias for real? *Journal of Applied Social Psychology*, *34*, 2069-2098.
- Eagly, A., & Chaiken, S. (1993). *The psychology of attitudes*. New York: Harcourt Brace.
- Eisenberg, T., Hannaford-Agor, P.L., Hans, V.P., Waters, N.L., Munsterman, G.T., Schwab, S.J., et al. (2005). Judge-jury agreement in criminal cases: A partial replication of Kalven and Zeisel's *The American Jury*. *Journal of Empirical Legal Studies*, *2*, 171-206.
- Foley, L.A., & Pigott, M.A. (1997). The influence of forepersons and nonforepersons on mock jury decisions. *American Journal of Forensic Psychology*, *15*, 5-17.
- Greene, E., Chopra, S.R., Kovera, M.B., Penrod, S.D., Rose, V.G., Schuller, R., & Studebaker, C.A. (2002). Jurors and juries: A review of the field. In J.R.P. Ogloff (Ed.), *Taking psychology and law into the twenty-first century* (pp. 225-284). New York: Kluwer Academic/Plenum Publishing.
- Hannaford-Agor, P.L., & Hans, V.P. (2003). Nullification at work? A glimpse from the National Center for State Courts study of hung juries. *Chicago-Kent Law Review*, *78*, 1249-1277.
- Hannaford-Agor, P.L., Hans, V.P., Mott, N.L., & Munsterman, G.T. (2002). *Are hung juries a problem?* Williamsburg, VA: National Center for State Courts, National Institute of Justice.
- Hans, V.P., Hannaford-Agor, P.L., Mott, N.L., & Munsterman, G.T. (2003). The hung jury: *The American Jury's* insights and contemporary understanding. *Criminal Law Bulletin*, *39*, 33-47.
- Hastie, R. (Ed.). (1993). Introduction. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 3-41). New York: Cambridge University Press.

- Heise, M. (2004). Criminal case complexity: An empirical perspective. *Journal of Empirical Legal Studies, 1*, 331-369.
- Heuer, L., & Penrod, S.D. (1994). Trial complexity: A field investigation of its meaning and its effect. *Law and Human Behavior, 18*, 29-52.
- Kalven, H., & Zeisel, H. (1966). *The American Jury*. Chicago: University of Chicago Press.
- Kerr, N.L. (1978). Severity of prescribed penalty and mock jurors' verdicts. *Journal of Personality and Social Psychology, 36*, 1431-1442.
- Kerr, N.L., Hymes, R.W., Anderson, A.B., & Weathers, J.E. (1995). Defendant-juror similarity and mock juror judgments. *Law and Human Behavior, 19*, 545-567.
- Kerr, N.L., MacCoun, R.J., & Kramer, G.P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review, 103*, 687-719.
- Kerr, N.L., Niedermeier, K.E., & Kaplan, M.F. (1999). Bias in jurors vs bias in juries: New evidence from the SDS perspective. *Organizational Behavior and Human Decision Processes, 80*, 70-86.
- Levett, L.M., Danielsen, E.M., Kovera, M.B., & Cutler, B.L. (2005). The psychology of jury and juror decision making. In N. Brewer, & K.D., Williams (Eds.), *Psychology and law: An empirical perspective*. (pp. 365-406). New York: Guilford Press.
- Marques, J.M. (1990). The black sheep effect: Outgroup homogeneity in social comparison settings. In D. Abrams & M. Hogg (Eds.), *Social Identity Theory: Constructive and critical advances* (pp. 131-151). London: Harvester Wheatsheaf.

- Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24*, 1315-1344.
- Mitchell, T.L., Haw, R.M., Pfeifer, J.E., & Meissner, C.A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior, 29*, 621-637.
- Mize, G.E., Hannaford-Agor, P., & Waters, N.L. (2007). *The state-of-the-states survey of jury improvement efforts: A compendium report*. Retrieved January 16, 2008, from http://www.ncsconline.org/D_Research/cjs/pdf/SOSCompendiumFinal.pdf.
- Myers, M.A. (1979). Rule departures and making law: Juries and their verdicts. *Law and Society Review, 13*, 781-797.
- Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 192-221). New York: Cambridge University Press.
- Petty, R.E., & Cacioppo, J.T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: William C. Brown.
- Petty, R.E., & Cacioppo, J.T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.
- Reskin, B.F., & Visser, C.A. (1986). The impacts of evidence and extralegal factors in jurors' decisions. *Law and Society Review, 20*, 423-438.
- Saks, M.J. (1997). What do jury experiments tell us about how juries (should) make decisions? *Southern California Interdisciplinary Law Journal, 6*, 1-53.

- Stebly, N.M., Besirevic, J., Fulero, S.M., & Jimenez-Lorente, B. (1999). The effects of pretrial publicity on juror verdicts: A meta-analytic review. *Law and Human Behavior, 23*, 219-235.
- Strodtbeck, F.L., & Lipinski, R.M. (1985). Becoming first among equals: Moral considerations in jury foreman selection. *Journal of Personality and Social Psychology, 49*, 927-936.
- Studebaker, C.A., & Penrod, S.D. (2005). Pretrial publicity and its influence on juror decision making. In N. Brewer, & K.D., Williams (Eds.), *Psychology and law: An empirical perspective*. (pp. 254-275). New York: Guilford Press.
- Sweeney, L.T., & Haney, C. (1992). The influence of race on sentencing: A meta-analytic review of experimental studies. *Behavioral Sciences & the Law, 10*, 179-195.
- Taylor, T. S., & Hosch, H. M. (2004). An examination of jury verdicts for evidence of a similarity-leniency effect, an out-group punitiveness effect or a black sheep effect. *Law & Human Behavior, 28*(5), 587-598.
- Visher, C.A. (1987). Juror decision making. *Law and Human Behavior, 11*, 1-17.
- Werner, C.M., Strube, M.J., Cole, A.M., & Kagehiro, D.K. (1985). The impact of case characteristics and prior jury experience on jury verdicts. *Journal of Applied Social Psychology, 15*, 409-427.

Table 1. Means, Standard Deviations, and Correlations with Jury Verdicts (Three Most Severe Charges)

Predictor Variables	N	Mean	SD	Zero-Order			Controlling for SOE				
				Verdict, Charge 1	Verdict, Charge 2	Verdict, Charge 3	Verdict, Charge 1	Verdict, Charge 2	Verdict, Charge 3		
				N=177	N=117	N=63	N=177	N=117	N=63		
State's SOE Rating, Charge 1	158	4.23	1.11	0.48**	--	--	--	--	--	--	--
State's SOE Rating, Charge 2	101	4.39	1.26	--	0.60**	--	--	--	--	--	--
State's SOE Rating, Charge 3	50	4.34	1.59	--	--	0.43**	--	--	--	--	--
State's SOE Rating, Overall	156	4.31	1.09	0.43**	0.53**	0.62**	0.02	0.13	0.64**	0.64**	0.64**
Prosecution Evidence Presented	88	4.32	2.06	0.39**	0.44**	0.40*	0.30**	0.35**	0.46*	0.46*	0.46*
Pros. Attorney Competence	179	3.92	0.73	0.35**	0.15	0.38**	0.27**	0.08	0.30*	0.30*	0.30*
Defense Attorney Competence	179	3.63	0.75	-0.16*	-0.13	0.06	-0.15	-0.10	0.06	0.06	0.06
Pretrial Publicity (Yes)	138	0.14	0.35	0.26**	0.18	0.17	0.22**	0.16	0.18	0.18	0.18
Trial Complexity	158	1.86	0.67	0.33**	0.26**	0.44**	0.34**	0.25*	0.49**	0.49**	0.49**
Charge 1 Severity	178	7.06	1.45	0.32**	--	--	0.30**	--	--	--	--
Charge 2 Severity	110	6.13	1.56	--	0.15	--	--	0.22*	--	--	--
Charge 3 Severity	51	5.61	1.56	--	--	0.20	--	--	0.04	0.04	0.04
Defendant Race (White)	131	0.73	0.44	0.05	0.11	-0.07	0.06	0.13	-0.07	-0.07	-0.07
Defendant Gender (Male)	131	0.94	0.24	-0.04	-0.06	0.09	-0.10	-0.17	-0.05	-0.05	-0.05
Defendant Likeableness	179	2.71	0.61	-0.05	-0.05	-0.11	-0.07	-0.14	-0.17	-0.17	-0.17
Defendant Attractiveness	179	2.69	0.58	-0.04	0.04	0.13	-0.08	-0.04	0.06	0.06	0.06
Defendant Testimony (Yes)	177	0.42	0.49	-0.11	-0.16	-0.26*	-0.08	-0.08	-0.17	-0.17	-0.17
Percentage Jury, White	179	0.84	0.13	0.04	0.13	0.08	0.04	0.05	-0.05	-0.05	-0.05
Percentage Jury, Male	179	0.45	0.16	0.08	0.08	0.17	0.05	0.01	0.27	0.27	0.27
Percentage Jury, White Male	179	0.44	0.17	0.08	0.07	0.17	0.04	-0.04	0.21	0.21	0.21
Percentage Jury, White Female	179	0.40	0.16	-0.05	0.03	-0.10	-0.01	0.08	-0.23	-0.23	-0.23
Percentage Jury, Black Male	179	0.04	0.06	0.02	0.07	0.06	0.05	0.19	0.13	0.13	0.13
Percentage Jury, Black Female	179	0.10	0.09	-0.14	-0.27**	-0.18	-0.12	-0.23*	-0.11	-0.11	-0.11
Jury-Defendant Racial Similarity	131	3.51	3.85	0.03	0.05	-0.12	0.06	0.12	-0.15	-0.15	-0.15
Jury-Defendant Gender Similarity	131	5.05	1.91	0.16	0.13	0.14	0.11	0.06	0.20	0.20	0.20
Jury-Def. Composite Similarity	131	1.60	2.30	0.04	0.03	-0.17	0.07	0.08	-0.22	-0.22	-0.22
Foreperson Race (White)	179	0.87	0.34	0.20**	0.04	0.04	0.20*	-0.01	0.03	0.03	0.03
Foreperson Gender (Male)	179	0.71	0.46	0.17*	0.04	0.22	0.16*	-0.07	0.21	0.21	0.21

Notes: * $p \leq .05$. ** $p \leq 0.01$. *N* indicates cases available for respective variable; pairwise sample sizes are less than or equal to the lower *n* in each pair. Parenthetical text indicates value coded as "1" for dummy-coded variables.

Table 2.

Multivariate Logistic Regression for Charge 1 Verdict.

Predictor Variables In Model	B	SE	Wald	p	OR	Acc.
<i>Step 0</i>						52.2%
Constant	0.09	.17	.27	.607	1.09	
<i>Step 1</i> [$\Delta\chi^2(1) = 34.23, p < .001$]						77.2%
Prosecution Strength of Evidence	1.05	.21	24.82	.000	2.85	
<i>Step 2</i> [$\Delta\chi^2(5) = 33.38, p = .001$]						80.1%
Prosecution Strength of Evidence	1.13	.25	21.04	.000	3.10	
Charge Severity	0.29	.19	2.39	.122	1.33	
Pretrial Publicity ^a	1.09	.84	1.69	.193	2.97	
Trial Complexity	0.97	.39	6.18	.013	2.64	
Foreperson Gender ^b	1.01	.54	3.55	.060	2.74	
Foreperson Race ^c	2.20	.94	5.45	.020	9.04	

Notes: N = 136 trials. B = regression coefficient. OR = change in odds ratio for one-unit

increment in predictor. Acc. = classification accuracy for model at a given step. ^a0=No, 1=Yes.

^b0=Female, 1=Male; ^c0=Black, 1=White.

Table 3.

Extraevidentiary-Verdict Correlations for Charge 1, by Evidentiary Category.

<i>Extraevidentiary Variable</i>	<i>Evidentiary Category</i>		
	Pro-Acquittal (SOE < 4.00) N = 45	Ambiguous (SOE 4.00-4.99) N = 52	Pro-Conviction (SOE > 4.99) N = 59
Charge 1 Severity	.22	.32*	.13
Pretrial Publicity ^a	.09	.39**	.22
Trial Complexity	.24	.54**	.17
Defendant Attractiveness	-.15	.06	-.07
Defendant Likeability	-.17	.17	-.08
Defendant Gender ^b	-.47**	.10	-.07
Defendant Race ^c	.21	-.01	.23
Foreperson Gender ^b	.30*	.18	.39**
Foreperson Race ^c	.07	.35*	.30*
% Jury Male	.07	.05	.16
% Jury White	-.02	.10	.15
% Jury, White Male	-.02	.03	.20
% Jury, White Female	.00	.03	-.09
% Jury, Black Male	.22	-.01	-.13
% Jury, Black Female	-.13	-.12	-.23
Jury-Defendant Gender Similarity	-.15	.12	.25
Jury-Defendant Race Similarity	.13	.06	.17
Jury-Defendant Composite Similarity	.02	.00	.24

Notes: *p ≤ .05. **p ≤ .01. ^a0=No, 1=Yes. ^b0=Female, 1=Male. ^c0=Non-White, 1=White.