

The Construction and Validation of a Juror Bias Scale

SAUL M. KASSIN

Williams College

AND

LAWRENCE S. WRIGHTSMAN

University of Kansas

A questionnaire was constructed to measure individual differences in pretrial bias among jurors. The final Likert scale, called the Juror Bias Scale (JBS), contains 17 items—8 that reflect pretrial expectancies that defendants, in general, commit the crimes with which they are charged and 9 that reflect the value attached to conviction and punishment. The scale is internally consistent and test-retest reliable. Scores are uncorrelated with social desirability, moderately correlated with I-E control and belief in a just world, and more highly correlated with authoritarianism. In one validation experiment, student jurors were exposed to three trial presentations in a laboratory setting. Overall, subjects classified as prosecution biased were more conviction prone and adopted a less stringent standard of reasonable doubt. In a second study, community jurors watched one of two videotaped mock trials in a courtroom. Prosecution-biased subjects asserted a higher probability that the defendant committed the crime and rendered a higher percentage of guilty verdicts than defense-biased subjects for one of the two trials. JBS scores were unrelated to all demographic variables, but were significantly correlated with political views. The potential uses and limitations of the JBS are discussed.

The American system of criminal justice was founded on the idea that an accused person must be tried by an impartial tribunal and solely on

This research was supported in part by University of Kansas General Research Grant 3401 to the second author and in part by Purdue University funds to the first author. We thank Raymond Stone, Keith Salette, Kevin Gibson, Lenny Spagnolo, Robert Fisher, Mike McClarney, Ani Rodriguez, and Elaine Rosenfeld for their role as experimenters. We would also like to thank John Meyers, Tippecanoe County prosecutor, for his help in arranging for subjects and facilities for Study II, Dave Freedman and Ed Arnolds for making available copies of their *Trial Diplomacy Journal* questionnaire, Laurence Rose and his University of Kansas law students for their reenactment of the Lynch trial, and Richard Smilgoff of the Court Practice Institute, Inc., for providing the Burks trial videotape. Reprint requests should be addressed to Saul Kassin at the Department of Psychology, Williams College, Bronfman Science Center, Williamstown, MA 01267.

the basis of evidence admitted in court. Despite the numerous mechanisms that have evolved for eliminating nonevidentiary sources of bias, legal scholars, practitioners, and researchers acknowledge in word and in deed that the ideal condition of pretrial neutrality is a seldom achieved "legal fiction" (Marshall, 1980). In fact, two interacting sources of pretrial bias can be distinguished—(a) relatively enduring personal characteristics of jurors, and (b) situational, case-specific influences such as media publicity (e.g., Padawer-Singer & Barton, 1975) and the demeanor of attorneys (Kaplan & Miller, 1978). This paper focuses on the former in an attempt to construct and validate an individual-difference measure of juror bias.

Personal Characteristics of Jurors—An Overview

Almost everyone in the legal community has an implicit or explicit theory about the match between types of jurors and their courtroom decisions. For years, trial lawyers have sought to identify the personality, attitudinal, and demographic characteristics that predict prospective jurors' inherent biases so as to challenge at voir dire those who will prove unfavorable to their side. As early as 1917, Brumbaugh wrote about several demographic biases. The current trial advocacy literature still contains a variety of intuitive heuristics (e.g., "cabinetmakers . . . should be avoided because they require everything in the case to fit together neatly," cf. Mossman, 1973). Indeed, empirical study of practicing attorneys' jury selection strategies confirms that they often do rely on judgmental heuristics of questionable validity and generality. Zeisel (1977) found that jurors' intelligence, age, occupation, physical appearance, and gender were among the most salient (and, of course, accessible) characteristics. Penrod (Note 1) found that the most frequently asked voir dire questions pertained to prospective jurors' attitudes about the particular crime and about the police.

The belief among lawyers that individual jurors may be prejudiced and that their prejudice can overwhelm the more evidentiary bases of a decision is widespread. However, attempts to identify broad demographic variables (e.g., race, sex, SES) and personality constructs (e.g., locus of control, belief in a just world) that consistently predict jurors' verdicts across trials have met with only limited success (Bridgeman & Marlowe, 1979; Davis, Bray, & Holt, 1977; Elwork, Sales, & Suggs, 1981; Gerbasi, Zuckerman, & Reis, 1977; Mills & Bohannon, 1980; Moran & Comfort, 1982; Saks & Hastie, 1978; Stephan, 1975).

To date, the most effective personality predictor of mock juror decisions has been authoritarianism (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) as measured by different versions of the California F-Scale (cf. Byrne, 1974). Not unexpectedly, authoritarians tend to be relatively punitive toward criminal defendants (for reviews, see Davis et al., 1977; Elwork & Sales, 1980). There are, however, important qualifications to

this pattern. First, although authoritarians are demonstrably harsher in their sentencing recommendations, there is little evidence that they are more conviction prone in their judgments of guilt (Bray & Noble, 1978, is an exception). Second, some research failed to support the punitiveness prediction (Gladstone, 1969; Sue, Smith, & Pedroza, 1975; Thayer, 1970), and other studies have even obtained the reverse pattern. Specifically, those who score as highly authoritarian are *less* punitive than those who score low on the F-Scale when the defendant is an authority figure like a policeman (Mitchell, Note 2) or when the crime being judged reflects obedience to (Hamilton, 1976) or the exercise of (Garcia & Griffitt, 1978) authority. Similar results are reported for the conceptually overlapping dogmatism construct (Shaffer & Case, 1982).

In contrast to the global construct approach, several investigators, in need of a juror predisposition measure to test juror decision models, have classified subjects on the basis of their attitudes in criminal/legal issues. Kaplan and Miller (1978) thus employed Wang and Thurstone's Attitude Toward Punishment of Criminals Scale, Boehm (1968) constructed an authoritarian legal attitudes questionnaire, and Marshall and Wise (1975) assessed attitudes toward the death penalty. Ostrom, Werner, and Saks (1978) measured attitudes in an even more relevant and focused domain—they combined subjects' Likert responses to five pro/antidefendant statements (e.g., "Most people who are brought to trial are guilty as charged") to form a measure of generalized orientation toward defendants. This latter strategy for isolating individual differences makes conceptual sense, and, indeed, this instrument did predict certain aspects of mock jurors' decisions. The validity of this 5-item instrument is questionable, however, for two reasons. First, subjects made their decisions on the basis of brief written trial summaries that differed from the full-blown trial in both amount of information and the mode through which it was presented. As several critics of the simulated jury paradigm have noted (e.g., Miller, Fontes, Boster, & Sunnefrank, Note 3) this truncated and oversimplified type of presentation may act to spuriously inflate the importance of an independent variable, such as pretrial bias. Second, Ostrom et al. (1978) had subjects rate the "probability of guilt," a scalar variable that represents only a partial determinant of the practically important measure—the true guilty vs. not guilty dichotomous verdict.

THE CONSTRUCTION OF A JUROR BIAS SCALE (JBS)

Rationale

Several theories of juror decision making incorporate a pretrial disposition component into their analyses of the judgmental process (cf. Kaplan, 1982). It is conceivable that systematic individual differences thus exist in the way jurors attend to, organize, and retrieve evidentiary information, in their assessments of witness's credibility, and in their evaluations of

the lawyers' arguments, in their interpretations of judges' instructions, and so on.

In order to test these effects of generalized pretrial bias, the forgoing literature review suggests the need for a reliable and valid self-report instrument designed specifically to assess peoples' predispositions as jurors toward guilt or innocence. The construction of such a test was guided by the fact that virtually all juror models assume that verdicts reflect the implicit operation of two "decisions": (a) probability of commission (PC), that is, the subjective likelihood (0–100%), given one's a priori beliefs and the evidence, that the defendant actually committed the crime; and (b) reasonable doubt (RD), that is, the threshold of certainty (0–100%) deemed necessary for conviction (cf. Pennington & Hastie, 1981, for a review of the various decision models). Thus, judgments of guilt arise when a juror's PC estimate exceeds his/her RD criterion, while the not-guilty verdict follows when a juror's PC estimate falls short of his/her "beyond a reasonable doubt" threshold. This analysis implies that individual jurors might differ from each other along two *theoretically* independent dimensions—generalized PC and generalized RD. Within this framework, scale items that reflect both components were generated.

Item Writing/Selection

As a first step, 43 statements were written (21 RD, 22 PC) and group administered in Likert format (where 1 = strongly agree, 2 = moderately agree, 3 = agree and disagree equally, 4 = moderately disagree, 5 = strongly disagree) to students at the University of Kansas ($n = 86$) and Purdue University ($n = 98$). Approximately half the items were worded so that an "agree" response indicated a prosecution bias, and half were so worded that "agree" indicated a defense bias.¹ Three criteria were adopted for inclusion of an item into the final scale: the item had to elicit a varied range of responses across the five Likert categories, it had to be significantly correlated with the total score (minus that item), and it had to have a relatively low correlation with scores on the Crowne and Marlowe (1964) social desirability scale that was administered after the juror bias items. After this first phase of testing, nine statements met the above criteria. A number of statements were thus rephrased and new items were written—30 in all were retested in a second sample of 107 Purdue students.

The final scale consists of 17 items that ultimately fulfilled the three criteria. Nine are statements designed to broadly reflect PC or subjective expectancy differences (e.g., "any suspect who runs from the police probably committed the crime," "circumstantial evidence is too weak to use in court") and eight were designed to measure the RD or "utility"

¹ This original pool of items is available upon request from the first author.

component (e.g., "too often jurors hesitate to convict someone who is guilty out of pure sympathy," "too many innocent people are wrongfully imprisoned"). As it turned out, the two sets of items were highly inter-correlated ($r(106) = .60, p < .001$) and so were combined to form a single unidimensional scale. An individual's total score is thus obtained by reversing his or her scores on all prosecution-worded statements (i.e., 5 = strongly agree . . . 1 = strongly disagree) and then summing across the 17 items. Scores could range from 17 to 85, where high numbers indicate a generalized prosecution (P) bias and low scores a defense (D) bias. Five filler items that pertain generally to legal issues are included in order to disguise somewhat the specific purpose of the scale, but do not enter into the scoring (e.g., "appointed judges are more competent than elected judges"). The entire scale, entitled "legal opinions survey," is presented in Table 1.

Test Characteristics

The final form of the scale (henceforth referred to as the JBS) was administered to three groups of introductory psychology majors (total $n = 221$). One group ($n = 101$) also completed additional personality and attitude scales that have previously been employed in jury research. Overall, JBS scores ranged from 39 to 66 with a M of 50.88 and a SD of 7.01. Internal consistency based on a split-half reliability of the 221 scores was .81. For 31 subjects who returned for a second questionnaire session 5 weeks later, the test-retest reliability was .67 ($p < .001$).

Correlations with other scales

The JBS was administered in conjunction with the Crowne-Marlowe (1964) social desirability scale, the internal-external locus of control scale (Rotter, 1966), the belief in a just world scale (Rubin & Peplau, 1975), the balanced F-Scale (Byrne, 1974), and the Thurstone attitudes toward capital punishment scale (cf. Shaw & Wright, 1964).

Intertest correlations showed first that the JBS was uncorrelated with social desirability ($r = -.01$), suggesting that subjects' responses reflect their position on the content of the items and not the operation of a self-presentation strategy. On the substantive scales, the JBS was moderately correlated with I-E ($r(100) = .23, p < .01$), moderately correlated with Just World ($r(100) = .24, p < .01$) and, not unexpectedly, highly correlated with authoritarianism ($r(100) = .43, p < .001$). These latter three scales, previously employed in mock juror research, were all significantly correlated with social desirability ($r_s = .18, .29, .24$, respectively), indicating that the JBS possesses a unique advantage as a measuring instrument. Surprisingly, the JBS was not significantly related to attitudes toward capital punishment ($r(30) = .21, p < .15$), though the latter correlation was based only on the relatively small sample of 31 retest subjects.

TABLE 1
ITEMS AND STATISTICS FOR THE JUROR BIAS SCALE^a

Items and scoring key ^b	Student sample (n = 101)			Community sample (n = 85)		
	M ^d	SD	r ^c	M	SD	r
(Filler)	—	—	—	—	—	—
1. Appointed judges are more competent than elected judges.	—	—	—	—	—	—
(P)	3.17	1.23	.53	3.47	1.17	.53
2. If a suspect runs from the police, then he probably committed the crime. (PC) ^c	—	—	—	—	—	—
(P)	2.51	1.28	.38	2.73	1.53	.55
3. A defendant should be found guilty if 11 out of 12 jurors vote guilty. (RD)	—	—	—	—	—	—
(Filler)	—	—	—	—	—	—
4. Most politicians are really as honest as humanly possible.	—	—	—	—	—	—
(P)	2.78	1.14	.44	3.26	1.21	.39
5. Too often jurors hesitate to convict someone who is guilty out of pure sympathy. (RD)	—	—	—	—	—	—
(P)	3.10	1.18	.45	3.01	1.19	.52
6. In most cases where the accused presents a strong defense, it is only because of a good lawyer. (PC)	—	—	—	—	—	—
(Filler)	—	—	—	—	—	—
7. In general, children should be excused for their misbehavior.	—	—	—	—	—	—
(D)	3.51	1.43	.48	3.94	1.33	.46
8. The death penalty is cruel and inhumane. (RD)	—	—	—	—	—	—
(P)	3.07	.96	.58	3.52	1.07	.51
9. Out of every 100 people brought to trial, at least 75 are guilty of the crime with which they are charged. (PC)	—	—	—	—	—	—
(P)	2.32	1.33	.50	2.15	1.37	.50
10. For serious crimes like murder, a defendant should be found guilty so long as there is a 90% chance that he committed the crime. (RD)	—	—	—	—	—	—
(P)	2.66	1.24	.46	2.58	1.39	.40
11. Defense lawyers don't really care about guilt or innocence, they are just in business to make money. (PC)	—	—	—	—	—	—
(P)	2.67	1.36	.51	3.00	1.34	.50
12. Generally, the police make an arrest only when they are sure about who committed the crime. (PC)	—	—	—	—	—	—

(D)	13. Circumstantial evidence is too weak to use in court. (PC)	2.73	1.23	.37	2.76	1.29	.19
(P)	14. Many accident claims filed against insurance companies are phony. (PC)	3.39	1.14	.45	3.65	1.12	.37
(D)	15. The defendant is often a victim of his own bad reputation. (PC)	2.49	1.00	.33	2.28	.97	.11
(P)	16. If the grand jury recommends that a person be brought to trial, then he probably committed the crime. (PC)	2.80	1.89	.48	2.81	1.17	.41
(P)	17. Extenuating circumstances should not be considered—if a person commits a crime, then that person should be punished. (RD)	3.11	1.36	.38	3.43	1.25	.43
(Filler)	18. Hypocrisy is on the increase in society.	—	—	—	—	—	—
(D)	19. Too many innocent people are wrongfully imprisoned. (RD)	3.01	1.14	.52	3.48	1.11	.33
(D)	20. If a majority of the evidence—but not all of it—suggests that the defendant committed the crime, the jury should vote <i>not guilty</i> . (RD)	3.08	1.22	.56	3.11	1.38	.48
(D)	21. If the defendant committed a victimless crime like gambling or possession of marijuana, he should never be convicted. (RD)	3.64	1.36	.40	4.07	1.17	.27
(Filler)	22. Some laws are made to be broken.	—	—	—	—	—	—

Note. Letters in parentheses to the left of each item indicate direction of wording, where P = prosecution, and D = defense.

^a Instructions to the legal opinions survey were: This is a questionnaire to determine the legal attitudes and beliefs of different people on a variety of statements. Please answer each statement by giving as true a picture of your own beliefs as possible. If you strongly agree with an item, mark the letter A. Mark the letter B if you mildly agree with the item—that is, if you think the item is generally more true than untrue. Mark the letter C if you feel the item is about equally true as untrue. Mark the letter D if you mildly disagree with the item—that is, if you think the item is generally more untrue than true. If you strongly disagree with the item, mark the letter E.

^b The scale is keyed in the direction of prosecution bias. For P statements, strongly agree = 5, strongly disagree = 1. For D statements, strongly agree = 1, strongly disagree = 5.

^c Items followed by a PC were designed to reflect the probability-of-commission or the subjective expectancy component; RD items pertain to the reasonable doubt or subjective value component.

^d Mean scores for each item can range from 1 to 5 and are based on the recorded (i.e., strongly agree with P statements and strongly disagree with D statements = 5) responses.

^e r represents the item-test (minus that item) correlation coefficients.

Preliminary Validation Study

As a first step, we sought to determine whether those subjects biased toward the defense differed from those biased toward the prosecution in their judgments of guilt based on their reading brief, relatively uninformative case summaries. Forty-nine undergraduates filled out the JBS and then received a packet containing four 2- to 3-page case summaries, each followed by a brief verdict questionnaire. The four cases involved an auto theft (Juhnke, Vought, Pyszczynski, Dane, Losure, & Wrightsman, 1979), murder (Jurow, 1971, Case 1), traffic felony (Kaplan & Kemmerick, 1974), and bribery (gleaned from a mock trial). Their order of presentation was fully counterbalanced, yielding 24 stimulus packets.

Subjects were classified as P ($n = 25$) or D ($n = 24$) biased via median split on the JBS (median = 51.50). The number of guilty verdicts was summed for each subject, yielding verdict scores that could range from 0 to 4. An analysis of these scores was statistically significant ($t(47) = 2.98$, $p < .01$)—P-biased subjects voted guilty more frequently ($M = 2.88$ for a 72% conviction rate) than did D-biased subjects ($M = 1.75$ for a 43.75% conviction rate). In short, the JBS scores predicted subjects' verdict tendencies under abbreviated and admittedly artificial trial conditions.

VALIDATION STUDY I

The main criterion by which to assess the validity of the JBS is unambiguous—individual jurors' predeliberation judgments of guilty or not guilty. Fishbein and Ajzen (1974) distinguished between different levels of behavioral criteria that can be employed—the single observation of a single act, repeated observations of a single act (under either homogeneous or heterogeneous conditions), and repeated observations of multiple, functionally related acts (see also Epstein, 1980). Because the JBS was constructed as a measure of generalized (i.e., across defendants, types of crime, and trials) pretrial bias, it should consistently predict individuals' mean verdict tendencies, though not necessarily their responses to any specific trial. Within Fishbein and Ajzen's (1974) framework, it is therefore clear that the most appropriate JBS validation strategy is to make repeated observations of a single act (verdict) under heterogeneous conditions (i.e., across different stimulus trials). In contrast to the preliminary study reported above, this research was conducted within a more realistic mock juror paradigm.

Method

Subjects and Design

Forty-eight introductory psychology majors were recruited from a mass testing session in which the JBS was administered along with the personality inventories noted earlier.

The study was advertised as a three-part experiment, so subjects were encouraged to volunteer *only* if they expected to attend all sessions. As it turned out, 39 students completed the entire experiment and comprised our final sample.

When all the data were collected, subjects were classified as P-biased ($n = 17$) or D-biased ($n = 22$) via median split of their JBS scores (median = 51.47).

Stimulus Trials

All subjects were exposed to three stimulus trials in partially counterbalanced order (123, 231, 312). Since the impact of pretrial bias and other psychological variables is limited to ambiguous (i.e., neither too strong nor too weak) cases, the following trials were selected and edited to elicit variability in judgments.

Trial 1. One trial, entitled "U.S. v. Ron Oliver," is a 1-hr, 10-min black-and-white videotape of an auto theft case (cf. Juhnke et al., 1979; Kassin & Wrightsman, 1979) that was reenacted in a courtroom by Washburn University law students and videotaped from a juror's perspective. Substantively, the trial was based on an actual criminal case in which the defendant, Ron Oliver, was charged with stealing a car and transporting it across state lines. The government's case was based on the testimony of a used car salesman who identified Ron Oliver as the person who stole the car from the lot, and the statement of a highway patrolman who stopped the defendant for speeding and subsequently made the arrest. The defendant, on the other hand, testified that he was driving an acquaintance's car and had no knowledge that the vehicle had been stolen. The entire trial presentation consisted of opening statements, the examination of three witnesses, and closing arguments. Previous research with this version of the tape had produced a .56 conviction rate (Kassin & Wrightsman, 1979, the single judgment-no instruction cell, $n = 18$).

Trial 2. The second trial, entitled "U.S. v. Lynch," was based on an actual conspiracy case. It was performed in a courtroom by University of Kansas law students and videotaped in black and white from the jury box. In this trial, the defendant—Bonnie Lynch—was charged with "willfully and knowingly harboring and concealing Frank Adams for whose arrest a federal warrant on a charge of felony had been issued." Specifically, the prosecution claimed that the defendant (a) accepted money from Frank Adams as payment for her protection, (b) lied to her landlord who inquired about the presence of a stranger in Bonnie Lynch's apartment, (c) drove Adams across state lines to a bus station, and (d) purchased a bus ticket for Adams. In support of these charges, the government introduced two witnesses. Jesse Nolan, a co-conspirator who was granted immunity for his testimony, testified that he made the arrangements for the defendant. Alma Richards, the defendant's landlady, testified that she saw Adams in the defendant's apartment and that the defendant behaved secretly, denying his presence. The defense claimed, essentially, that Bonnie Lynch was not aware of Adams' record, having been misinformed by Jesse Nolan. The defendant and her brother, who was present when Adams first arrived, both testified in support of this claim.

The original 1-hr, 40-min videotape yielded only a .20 conviction rate in pretesting ($n = 20$). Several defense-oriented testimony/arguments were thus deleted. This edited version, approximately 1 hr and 10 min in length, elicited a more suitable .44 conviction rate in pretesting ($n = 16$). It included opening statements, the examination of four witnesses, closing arguments, and instructions from the judge.

Trial 3. The third presentation consisted of an 18- to 19-page adaptation of the Adams-Zemp assault case originally created by Walker, Thibaut, and Andreoli (1972). The transcript was written with the prescaled facts provided by Walker et al. (1972) and presented as a criminal trial entitled "Illinois v. Adams." In this case, Samuel Adams was charged with assault for stabbing and seriously injuring Michael Zemp with a piece of broken glass during a heated argument in a tavern. The defense claimed that Adams, feeling threatened and endangered, had acted in self-defense. The entire transcript contained opening remarks,

the examination of seven witnesses (including the defendant and the victim), and the judge's instruction. The version of the transcript that was chosen for the present study (others were written and tested) yielded a .60 conviction rate in pretesting.

Procedure.

Subjects were scheduled for one of three trial presentations. Two weeks after the JBS was administered, they participated in their first experimental session whereupon they were scheduled for a second session, etc. Each trial presentation was conducted by a different experimenter in small group settings ($n = 3$ to 5 subjects per session). Sessions were separated by 1- to 2-week intervals.

Upon entering each session, subjects were instructed as follows: "This study is part of an ongoing project on the decision making process of jurors. You will watch an edited videotape (read an edited transcript) of a criminal trial entitled _____. Please pay close attention (read carefully) and do not talk to each other during the trial. Afterwards, you will be asked to play the role of jurors, render a verdict, and answer other case-related questions."

In each instance, subjects then watched the trial and filled out a two-page questionnaire individually and without deliberation. Subjects received experimental credit for their participation after each session but were fully debriefed in writing about the experiment only after all the data were collected.

Dependent Measures

The questionnaire format was identical for the three trials. First, subjects rendered a dichotomous judgment (guilty-not guilty) and indicated their confidence in that verdict on a 0-8 scale. Second, they provided a quantitative, case-specific definition of reasonable doubt by filling in: "The defendant should be found guilty if there is at least a ____% chance that he/she committed the crime." Third, they indicated the probability that the defendant committed the crime by circling a number from 0 to 100 scaled in multiples of 5. Fourth, subjects rated the extent to which their decision was influenced by (a) each of the witnesses' testimony, and (b) each of the attorneys' arguments. All ratings were made on 0- to 8-point scales. Finally, all subjects were told to "assume the defendant had been convicted. If you were the judge, what kind of sentence would you recommend?" (where 0 = minimum allowed by law, 8 = maximum allowed by law).

Results and Discussion

In order to achieve a repeated observations criterion, the data from the three trials were combined by computing the means for identical response measures. Since each trial involved different numbers of witnesses, the ratings for all prosecution witnesses and attorneys were combined to form an overall measure of how effective the governments' cases were. The ratings for all defense witnesses and attorneys were similarly combined to form a measure of the defenses' effectiveness.

The three trials elicited a combined conviction rate of .615. For analyses, each subject was assigned a verdict score of 0-3 which represented his or her total number of guilty votes. As predicted, subjects who were classified by the JBS as P-biased voted guilty significantly more frequently ($M = 2.24$) than did those whose scores indicated a defense bias ($M = 1.55$). Put another way, the combined conviction rates were .745 and .515 for the P- and D-biased subjects, respectively ($t(37) = 2.62, p <$

.02). The two groups did not significantly differ in their mean level of verdict confidence.²

An analysis of subjects' case-specific RD and PC estimates revealed, quite unambiguously, the source of their verdict differences. That is, the two bias groups essentially agreed on the combined probability that the three defendants had committed the crimes with which they were charged (M 's = 73.67 and 74.77, $t(37) = -.19$, n.s.), but differed significantly in their interpretations of "beyond a reasonable doubt" ($t(37) = -2.45$, $p < .02$). P-biased subjects stated a mean willingness to vote guilty on the basis of an 83.23% certainty, whereas D-biased subjects set a more stringent standard of proof—91.47%. The mean correlation between within-trial estimates of PC and RD ($r = .15$) was nonsignificant, suggesting that the two decisions, when made in reference to a specific case, were orthogonal.³

Although P-biased subjects rated the government witnesses/attorneys across the three trials as collectively more influential than did D-biased subjects (M 's = 4.92 and 4.48 on a 0–8 scale), this difference was not statistically significant ($t(37) = 1.33$, $p < .20$). The two groups did not significantly differ in their overall ratings of the defense either (M 's = 4.64 and 4.88, $t(37) = -.64$). Prosecution-biased subjects recommended somewhat harsher sentences ($M = 4.35$) than did D-biased subjects ($M = 3.56$), but this difference only approached significance ($t(37) = 1.68$, $p < .10$).

Finally, how did the JBS compare to other personality measures for which subjects' scores were available? The correlations between each score and the verdict measure were $r = .22$ ($p < .10$) for Just World, $r = -.04$ ($p < .50$) for internal–external control, $r = .17$ ($p < .20$) for social desirability, and $r = .28$ ($p < .05$) for authoritarianism ($r = .37$ for the JBS, $p < .01$).

In sum, the JBS successfully predicted the verdicts of mock jurors, with P-biased individuals exhibiting a higher conviction rate than D-biased subjects. Moreover, the difference appeared to reflect the fact that although both groups derived from the evidence the same subjective likelihood that the defendants committed their crimes, the D-biased subjects set a higher standard of reasonable doubt, demanding greater certainty as necessary for conviction.

² An analysis of each trial separately revealed that for P- and D-biased subjects, respectively, the conviction rates were .88 and .54 for Trial 1 ($\chi^2(1) = 5.05$, $p < .03$), .59 and .36 for Trial 2 ($\chi^2(1) = 1.95$, $p < .15$), and .76 and .64 for Trial 3 ($\chi^2(1) < 1$, n.s.).

³ As with verdicts, the PC and RD estimates were analyzed separately for each trial. It turned out that P- and D-biased subjects did not differ in their PC estimates for any trial. For RD, however, P- and D-biased subjects, respectively, provided standards of 84 and 92% for Trial 1 ($t(37) = -1.91$, $p < .06$), 85 and 92% for Trial 2 ($t(37) = -1.73$, $p < .09$), and 80 and 91% for Trial 3 ($t(37) = -2.14$, $p < .04$).

VALIDATION STUDY II

The previous experiment established the predictive utility of the JBS vis-à-vis the criterion of mock jurors' repeated (i.e., across trials) verdicts. A second validation study was conducted with three goals in mind—(a) to increase the experimental and mundane realism of our mock jury paradigm by conducting the trial presentation in a real courtroom with subjects who expect to deliberate as a jury, (b) to investigate the demographic characteristics of P- and D-biased jurors by testing a heterogeneous sample of community residents selected from an actual jury list, and (c) to examine more closely how P- and D-biased subjects perceive various witnesses/testimony and attorneys/arguments. Two stimulus trials were employed, but each subject participated in only one—hence, a single-act criterion.

Method

Subjects

Eighty-five residents of Lafayette, Indiana, participated in this study—53 observed one trial and 32 the other. Subjects' names were taken from the 1979–1980 Tippecanoe County jury lists. They were called by one of two experimenters and offered \$6 to participate for 2–3 hrs in a practice jury.

Approximately 250 prospective subjects were contacted to obtain the 85 subjects, so a self-selection problem precluded our achieving a truly representative sample. Nevertheless, the variability among subjects was sufficient for our test of individual differences. Demographically, the 85 participants possessed the following characteristics: 43 female, 42 male; mean age = 36.91 (range from 18 to 69); mean education level = 13.19 grades (range from 9th grade to Ph.D.); mean income range = \$18,000–\$19,999.

Overview of Procedure

The procedure was identical for the two trials. Subjects who consented to participate appeared in groups ranging in size from 4 to 13. All subjects within a given session watched the stimulus trial together. Groups of 4–7 then deliberated as a single jury, whereas groups of 8–13 were divided for deliberation into two juries. The study was conducted during the evenings in a courtroom in Lafayette City Hall.

Upon entering, subjects were seated together in a "jury box" and asked to fill out the JBS. When all had completed it, the experimenters, two male undergraduates, briefed them about the videotaped trial they would see and instructed them about their role as practice jurors. At that point, either the *People v. Burks* rape trial or the *U.S. v. Lynch* conspiracy trial was shown on a 25-in videotape monitor that was placed on the judge's bench. When the trial presentation ended, subjects filled out a 3- to 4-page predeliberation questionnaire on which they rendered a verdict and answered a series of other case-specific questions. Next, subjects were escorted to their deliberation rooms and given 30 min to elect a foreperson, discuss the evidence, and arrive at a unanimous verdict.⁴ Finally, all subjects completed an extensive postdeliberation questionnaire which contained a wide variety of questions about demographic, personal, and experiential characteristics typically thought to be related to jurors' verdicts. Included were questions about sex, age, income,

⁴ Because these group data are not available for the present purposes, they will not be discussed further.

education, religion, religiosity, political views, and experience with crime/law enforcement officials.

To summarize, the procedure consisted of five phases: administration of the JBS, the trial presentation, the dependent variables questionnaire, jury deliberation, and the post-deliberation personal background questionnaire. At the conclusion of each session, subjects were debriefed fully about the nature of the experiment and paid for their participation.

Videotaped Stimulus Trials

U.S. v. Lynch. The first videotape was the mock conspiracy trial employed in Study 1—it had elicited a 46–54% split in verdicts.

People v. Burks. This black-and-white videotape is of a mock rape trial that was performed in a courtroom in the presence of a judge, an audience, and a mock jury, by nationally prominent lawyers (i.e., James A. Lindmark, prosecuting attorney; Julius L. Echeles, defense attorney).⁵ In this case, Louella Wilson, the complaining witness, testified that she was raped at gunpoint in the hallway of her apartment building but that the act was terminated prematurely when footsteps were heard. On the following day, she saw Herman Burks walking through the street, called the police, and identified him as her assailant. Herman Burks denied the charges, maintaining that he spent the entire day drinking at a friend's apartment in the victim's building (Louella Wilson had testified that she did not smell alcohol on the rapist's breath). The defense argued that Burks had an alibi and that the lighting conditions in the hallway were too poor for the victim to make an accurate identification. In addition, Louella Wilson's character and reputation for veracity were challenged.

The videotape consists of opening statements, the examination of seven witnesses, and closing arguments. In support of their case, the prosecution introduced three witnesses: the victim and two police officers who arrested Burks and testified that he had lied to them about his whereabouts. The defense called four witnesses: the defendant who denied the allegations, the defendant's mother and a friend who testified essentially about his character, and the defendant's friend/victim's neighbor who testified that he and Burks were drinking together on the day in question. The entire trial presentation is approximately 1 hr and 40 min in length.

Dependent Measures

The form of the predeliberation questionnaire was essentially the same for the Burks and Lynch trials. In both, subjects first rendered a verdict (guilty–not guilty), indicated their confidence (0–8) in that decision, quantified their definition of reasonable doubt (0–100), and rated their subjective likelihood (0–100) that the defendant committed the crime. Next, subjects rated all the witnesses and their testimony on three specific dimensions (cf. Morrill, 1972; Wigmore, 1937)—relevance, believability, and likability. They then rated the lawyers for each side on five characteristics—competent, prepared, sincere, likable, and persuasive. Finally, subjects were asked to suppose the defendant was convicted, and recommend a sentence. As in Experiment 1, all ratings were made on a 9-point (0–8) scale.

In addition to the above response measures, subjects who watched the rape trial indicated their agreement/disagreement with seven critical arguments in the case. The number of defense arguments with which each subject agreed (0–3) was subtracted from the number of prosecution arguments (0–4) to form a single measure.

⁵ This videotaped trial was performed for the Court Practice Institute, Inc., which generously provided it to the first author.

Results and Discussion

Lynch Conspiracy Trial

As before, subjects were classified as P-biased ($n = 17$) or D-biased ($n = 15$) by a median split (median = 56.32). Overall, 23 out of 32 subjects voted guilty, yielding a .72 conviction rate. This contrasts sharply with the .42 conviction rate obtained in the student sample of Experiment 1 and is consistent with the "leniency bias" previously reported for student (vs community) samples (Miller et al., Note 3).

For the Lynch conspiracy trial, P-biased subjects rendered significantly more guilty verdicts than D-biased subjects (88.24 and 53.33%, $\chi^2(1) = 4.80$, $p < .05$). In contrast to Study I, the two groups did not differ in their RD estimates (M 's = 87.53 and 87.87, $t(30) < 1$). The P-biased subjects did, however, indicate a higher PC than did D-biased subjects (M 's = 86.76 and 60.00, respectively; $t(30) = 2.63$, $p < .01$). No differences emerged for any of the witness or attorney ratings or for sentence recommendations. As in the first study, the correlation between PC and RD estimates was not significant ($r(30) = .14$).

Burks Rape Trial

Mock jurors were classified as P- and D-biased on the basis of a median split of their JBS scores (sample median = 52.13, yielding n 's of 25 and 28, respectively). Overall, 27 out of the 53 subjects voted guilty, producing a conviction rate of .51.

The two groups did not significantly differ in their verdicts. In fact, there was a nonsignificant reversal of the expected pattern ($\chi^2(1) = 2.27$, $p < .15$), as P-biased jurors were somewhat less likely to vote guilty than were D-biased jurors (40 and 60.71%, respectively). Consistent with these judgment data, the two groups did not differ in their RD or PC estimates ($t(51) = .03$ and -1.10 , respectively). In contrast to Study I, these latter estimates were significantly correlated with each other ($r(51) = .29$, $p < .05$).

Recall that subjects rated each witness on three characteristics and each lawyer on five. The sets of dimensions were highly intercorrelated (r 's ranged from .49 to .78), so they were summed for each witness (0–24) and lawyer (0–40). Analyses of these composite witness- and attorney-evaluation scores produced internally inconsistent results—D-biased jurors rated the defendant more favorably ($t(51) = -2.31$, $p < .025$) but also tended to rate the victim ($t(51) = -1.75$, $p < .10$) and another prosecution witness, a police officer, more favorably as well ($t(51) = -1.93$, $p < .06$).⁶ Finally, P- and D-biased subjects did not differ in their overall direction of agreement with the lawyers' arguments.

⁶ Interestingly, an inspection of *all* ratings (21 for witnesses, 10 for attorneys) reveals a remarkably consistent pattern—the D-biased subjects gave more favorable ratings on 30 out of 31 occasions (overall M 's = 5.10 and 4.47). This may reflect a positive response set on the part of the defense-biased subjects.

JBS Scores and Demographic Characteristics

Collapsed across the two trials ($n = 85$), the JBS elicited a mean score of 53.26 with a standard deviation of 8.79. These scores proved to be unrelated to mock jurors' sex, age, income, education, or religiosity. People who had previously served on a jury tended to score lower (i.e., defense biased, $M = 51.04$, $n = 21$) than those who did not have prior experience ($M = 54.25$), though this difference was not statistically significant ($t(81) = 1.37$, $p < .20$). Not surprisingly, one difference to emerge here involved subjects' self-reported position on the political spectrum ($F(2, 76) = 3.26$, $p < .05$)—self-described liberals ($M = 48.13$, $n = 16$) scored as more D-biased than did either moderates ($M = 55.04$, $n = 27$) or conservatives ($M = 53.82$, $n = 39$). Also, subjects who had been the victim of a *violent* crime or who had a close friend or relative who had been such a victim ($n = 20$) were more P-biased than those who had not had such an experience (M 's = 57.66 and 51.91, $t(83) = 2.01$, $p < .05$).⁷

GENERAL DISCUSSION

Overall, the present research has shown the JBS to be a generally reliable and valid self-report instrument for measuring individual differences among jurors. Specifically, JBS scores predicted (a) student jurors' verdict preferences for case summaries, (b) student jurors' verdict preferences and estimates of reasonable doubt for more extensive (videotaped and written) trial presentations, and (c) community jurors' verdict preferences and estimates of the probability of commission for one of two videotaped trial presentations. Moreover, the JBS—in contrast to the other previously employed scales (i.e., Just World, I-E, and authoritarianism)—was more highly correlated with verdict preferences (of the others, only the F-scale was significant) and was uncorrelated with social desirability. Collectively, these results suggest that the JBS has predictive utility that is unmatched by other, less focused instruments. Still, several important questions remain unanswered.

First, what is the domain of power of the JBS and what are its predictive boundaries? One obvious limitation, reflected in our choice and editing of stimuli, pertains to the strength of the evidence. As most psychologists in general and jury researchers in particular will agree, an individual's pretrial disposition will influence his or her decision only when situational cues are weak or ambiguous—as in a close, evenly balanced trial. A second limitation revealed itself in Study II where the JBS was effective for the conspiracy case but not for the rape trial which, in fact, produced a nonsignificant reversal of the expected P- and D-bias pattern. Why

⁷ No differences appeared for those who had (and had not) been the victim of a nonviolent crime ($n = 41$).

should the JBS predict verdict preferences for one ambiguous case but not another? At this point, we can only speculate since the two trials differed along several dimensions such as length of presentation, violence of crime, and emotional impact and tone of the testimony. Thus perhaps pretrial bias effects dissipate with increasingly long and detailed trials; or, perhaps the sympathies aroused by the emotional testimony of specific characters (e.g., a victim or a defendant) take precedence over relatively cognitive predispositions. Even more difficult to explain is the nonsignificant reversal obtained for the rape trial. Apparently, D-biased jurors, shown to be relatively liberal in their political views, are generally prone to favor the *victim* of a sexual assault crime. This possibility is supported by the correlation between subjects' JBS scores and their Likert responses to "Society's attitude toward sex is too permissive," an item contained in the postdeliberation questionnaire of study 2 ($r(84) = .30, p < .01$, where P-bias was associated with item endorsement) and is consistent with a $P \times S$ (juror \times trial) interactionist perspective (cf. Hans & Vidmar, 1982).

In order to test the above hypothesis more directly, the following additional data were collected: Sixty-two Williams College undergraduates (32 male, 30 female) were administered the JBS along with a recently published Rape Empathy Scale (RES) designed to measure peoples' empathy toward the victim versus the rapist in heterosexual assault situations (Deitz, Blackwell, Daley, & Bentley, 1982). Consistent with the results from the Burks trial of Study II, a significant correlation was obtained ($r = -.24, p < .05$), suggesting that P-biased jurors are generally *less* sympathetic to rape victims (and, conversely, are more sympathetic to rapists) than are D-biased jurors. Interestingly, further analysis revealed that this relationship held strongly for males ($r = -.44, p < .005$) but not for females ($r = .05, n.s.$).⁸

A second question raised by the present research is, how do juror predispositions, as measured by the JBS, operate? In Study I, prosecution- and defense-biased subjects differed in their criteria of reasonable doubt but not in their probability-of-commission estimates. In Study II, however, they differed in the latter but not in the former. This disparity is difficult to explain in view of differences between subject samples (student vs community), physical setting (laboratory vs courtroom), and task expectations (questionnaire only vs deliberation). At the very least, it suggests that the bias may exert its influence through either or both components. Alternatively, Nagel (1979) has suggested that pretrial bias produces a simple verdict preference and that jurors subsequently (e.g., in deliberation)

⁸ As in our earlier research, there were no sex differences in JBS scores (Males = 48.03, Females = 48.31). As reported by Deitz et al. (1982), however, females scored significantly higher than males on the RES (M 's = 117.37 and 103.03, respectively, $p < .001$).

alter and manipulate their PC estimates and RD values in order to justify that preference.

Having demonstrated the predictive validity of the JBS, it is important to delineate its limitations as well as its potential uses. First, the JBS was not designed as a jury selection instrument.⁹ Items for the scale were not selected on an empirical basis (i.e., by their relationships to a verdict criterion) but within a more substantive framework that emphasizes the conceptual relationship between a test item and its referent (cf. Loevinger, 1957; Jackson, 1971). As such, the JBS items offer little disguise of purpose and can be "faked" by an individual for whom jury service is either highly attractive or highly aversive. Additionally, the scale was constructed as a measure of generalized rather than case-specific predisposition. As such, it can be expected to provide a moderate level of prediction across a broad range of trials but an insufficient level of prediction for practical courtroom use in specific cases.

The JBS was constructed as a tool through which psychological models of the juror decision-making process, their personal, pretrial bias component, and the cognitive processes that sustain the latter, can be investigated. Implicit in the finding that pretrial beliefs and values prejudice verdict preferences is that they somehow overwhelm the "objective" and often conflicting evidence presented in court. What remains to be seen is whether P- and D-biased jurors differentially (i.e., in a schema-consistent manner) seek, attend to, organize, interpret, and/or recall the testimony, arguments, and instructions in a trial. Pennington and Hastie (1981) recently prescribed an idealized model of the juror decision-making process, outlining the tasks that jurors are successively confronted with during the course of a trial. Viewed within their framework, pretrial bias may affect a variety of predecision stages such as the selection of admissible evidence, the construction of a plausible sequence of events, the evaluation of credibility and probative value (e.g., of eyewitness testimony), and the application of the requirements of proof. As an individual differences

⁹ There are at least three ways in which the JBS might be so employed, though each has serious restrictions associated with it. The first strategy would be to administer the scale to an entire panel of prospective jurors, score their responses, and challenge those persons who emerge as the most unfavorably biased. One obvious and perhaps insurmountable problem would be eliciting the judge's approval for handing out a written questionnaire. A second strategy would be to choose "key items" from the scale, rephrase them in question-answer format, and incorporate them into the voir dire. The efficacy of this approach, however, rests on the validity of a small subset of key items. Moreover, unforeseeable difficulties may arise from the simple change in response mode and format of the items. A third approach is to administer the scale as part of a community-wide survey, compute the correlations between demographic variables and generalized bias, and peremptorily challenge venirepersons whose profiles suggest that they would assume an unfavorable position. The problem here is that if one were to conduct a survey, he or she would be better advised to construct more case-specific attitude measure (Christie, 1976) rather than employ the generalized JBS as a criterion.

test, the JBS could be fruitfully employed to elucidate such bias-sustaining processes.

REFERENCES

- Adorno, T. W., Frenkel-Brunswick, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
- Boehm, V. Mr. Prejudice, Miss Sympathy, and the authoritarian personality: An application of psychological measuring techniques to the problem of jury bias. *Wisconsin Law Review*, 1968, 734-750.
- Bray, R. M., & Noble, A. M. Authoritarianism and decisions of mock juries: Evidence of jury bias and group polarization. *Journal of Personality and Social Psychology*, 1978, 36, 1424-1430.
- Bridgeman, D. L., & Marlowe, D. Jury decision making: An empirical study based on actual felony trials. *Journal of Applied Psychology*, 1979, 64, 91-98.
- Brumbaugh, L. J. *Legal reasoning and briefing*. Indianapolis: Bobbs-Merrill, 1917.
- Byrne, D. *An introduction to personality*. Englewood Cliffs, New Jersey: Prentice-Hall, 1974.
- Christie, R. Probability v. precedence: The social psychology of jury selection. In G. Bermant, C. Nemeth, & N. Vidmar (Eds.), *Psychology and the law*. Lexington, Massachusetts: D. C. Heath, 1976.
- Crowne, D. P., & Marlowe, D. A. *The approval motive*. New York: Wiley, 1964.
- Davis, J. H., Bray, R. M., & Holt, R. W. The empirical study of social decision processes in juries. In J. Tapp & F. Levine (Eds.), *Law, justice, and the individual in society: Psychological and legal issues*. New York: Holt, Rinehart & Winston, 1977.
- Deitz, S. R., Blackwell, K. T., Daley, P. C., & Bentley, B. J. Measurement of empathy toward rape victims and rapists. *Journal of Personality and Social Psychology*, 1982, 43, 372-384.
- Elwork, A., & Sales, B. D. Psycholegal research on the jury and trial process. In C. Petty, W. Curran, & L. McGarry (Eds.), *Modern legal medicine and forensic science*. Philadelphia: Davis, 1980.
- Elwork, A., Sales, B. D., & Suggs, D. The trial: A research review. In B. Sales (Ed.), *The trial process*. New York: Plenum, 1981.
- Epstein, S. The stability of behavior: Implications for psychological research. *American Psychologist*, 1980, 35, 790-806.
- Fishbein, M., & Ajzen, I. Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 1974, 81, 59-74.
- Garcia, L. T., & Griffitt, W. Authoritarianism-situation interactions in the determination of punitiveness: Engaging authoritarian ideology. *Journal of Research in Personality*, 1978, 12, 469-478.
- Gerbasi, K. C., Zuckerman, M., & Reis, H. T. Justice needs a new blindfold: A review of mock jury research. *Psychological Bulletin*, 1977, 84, 323-345.
- Gladstone, R. Authoritarianism, social status, transgression, and punitiveness. *Proceedings of the 77th Annual Convention of the American Psychological Association*, 1969, 4, 287-288.
- Hamilton, V. L. Individual differences in ascriptions of responsibility, guilt, and appropriate punishment. In G. Bermant, C. Nemeth, & N. Vidmar (Eds.), *Psychology and the law*. Lexington, Massachusetts: Heath, 1976.
- Hans, V. P., & Vidmar, N. Jury selection. In N. L. Kerr & R. B. Bray (Eds.), *The psychology of the courtroom*. New York: Academic Press, 1982.
- Jackson, D. N. The dynamics of structured personality tests. *Psychological Review*, 1971, 78, 229-248.
- Juhnke, R., Vought, C., Pysczynski, T. A., Dane, F. C., Losure, B. D., & Wrightsman,

- L. S. Effects of presentation mode upon mock jurors' reactions to a trial. *Personality and Social Psychology Bulletin*, 1979, **5**, 36-39.
- Juror, G. Y. New data on the effect of a death qualified jury on the guilt determination process. *Harvard Law Review*, 1971, **84**, 567-611.
- Kaplan, M. F. Cognitive processes in the individual juror. In N. L. Kerr & R. M. Bray (Eds.), *The psychology of the courtroom*. New York: Academic Press, 1982.
- Kaplan, M. F., & Kemmerick, G. D. Juror judgment as information integration: Combining evidential and nonevidential information. *Journal of Personality and Social Psychology*, 1974, **30**, 493-499.
- Kaplan, M. F., & Miller, L. E. Reducing the effects of juror bias. *Journal of Personality and Social Psychology*, 1978, **36**, 1443-1455.
- Kassin, S. M., & Wrightsman, L. S. On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology*, 1979, **37**, 1877-1887.
- Loevinger, J. Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, **3**, 635-694 (Monograph No. 9).
- Marshall, J. *Law and psychology in conflict*. New York: Bobbs-Merrill, 1980. 2nd ed.
- Marshall, C. R., & Wise, J. A. Juror decisions and the determination of guilt in capital punishment cases: A Bayesian perspective. In D. Wendt & C. Vlek (Eds.), *Utility, probability, and human decision making*. Dordrecht, Holland: Reidel, 1975.
- Mills, C. J., & Bohannon, W. E. Juror characteristics: To what extent are they related to jury verdicts? *Judicature*, 1980, **64**, 22-31.
- Moran, G., & Comfort, J. C. Scientific juror selection: Sex as a moderator of demographic and personality predictors of impaneled felony juror behavior. *Journal of Personality and Social Psychology*, 1982, **43**, 1052-1063.
- Morrill, A. E. *Trial diplomacy*. Chicago: Court Practice Institute, 1972.
- Mossman, K. Jury selection: An expert's view. *Psychology Today*, 1973, May, 78-89.
- Nagel, S. Bringing the values of jurors in line with the law. *Judicature*, 1979, **63**, 189-195.
- Ostrom, T. M., Werner, C., & Saks, M. J. An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, 1978, **36**, 346-450.
- Padawer-Singer, A. M., & Barton, A. H. The impact of pretrial publicity on jurors' verdicts. In R. J. Simon (Ed.), *The jury system in America: A critical overview*. Beverly Hills, California: Sage, 1975.
- Pennington, N., & Hastie, R. Juror decision-making models: The generalization gap. *Psychological Bulletin*, 1981, **89**, 246-287.
- Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, **80**, (Whole No. 609).
- Rubin, Z., & Peplau, A. Who believes in a just world? *Journal of Social Issues*, 1975, **31**, 65-90.
- Saks, M. J., & Hastie, R. *Social psychology in court*. New York: Van Nostrand-Reinhold, 1978.
- Shaw, M. E., & Wright, J. M. *Scales for the measurement of attitudes*. New York: McGraw-Hill, 1964.
- Shaffer, D. R., & Case, T. On the decision to testify in one's own behalf: Effects of withheld evidence, defendant's sexual preferences, and juror dogmatism on juridic decisions. *Journal of Personality and Social Psychology*, 1982, **42**, 335-346.
- Stephan, C. Selective characteristics of jurors and litigants: Their influence on juries' verdicts. In R. J. Simon (Ed.), *The jury system in America: A critical overview*. Beverly Hills, California: Sage, 1975.
- Sue, S., Smith, R. E., & Pedroza, G. Authoritarianism, pretrial publicity and awareness of bias in simulated jurors. *Psychological Reports*, 1975, **37**, 1299-1302.

- Thayer, R. Attitude and personality differences between potential jurors who could return a death verdict and those who could not. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, 5, 445-446.
- Walker, L., Thibaut, J., & Andreoli, V. Order of presentation at trial. *Yale Law Journal*, 1977, 82, 216-226.
- Wigmore, J. H. *The science of judicial proof*. Boston: Little, Brown, 1937.
- Zeisel, H. The American Jury. In R. S. Jacobson (Ed.), *The American jury system: Final report*. Roscoe Pound-American Trial Lawyers Foundation, 1977.

REFERENCE NOTES

1. Penrod, S. *Practicing attorneys' jury selection strategies*. Paper presented at the Annual Meeting of the American Psychological Association, Montreal, September 1980.
2. Mitchell, H. E. *Authoritarian punitiveness in simulated juror decision making: The good guys don't always wear white hats*. Paper presented at the Annual Meeting of the Midwestern Psychological Association, Chicago, May 1973.
3. Miller, G. R., Fontes, N. E., Boster, F. J., & Sunnefrank, M. *Methodological issues in jury research: What can simulations tell us?* Paper presented at the Annual Meeting of the American Psychological Association, San Francisco, August 1977.