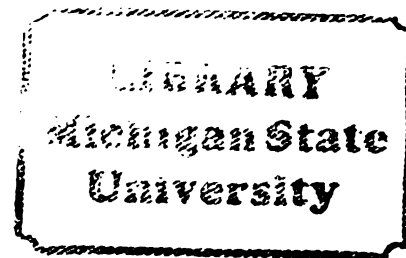THESIS

This is to certify that the

dissertation entitled

Modeling the Impact of Extralegal Bias and
Defined Standards of Proof on the Decisions of Mock
Jurors and Juries

presented by

Robert J. MacCoun

has been accepted towards fulfillment
of the requirements for

___Ph.D.___ degree in _Psychology_

Norbert L. Kerr
Major professor

Norbert L. Kerr

Date_August 16, 1984_

O-12771

MODELING THE IMPACT OF EXTRALEGAL BIAS AND DEFINED

STANDARDS OF PROOF ON THE DECISIONS OF MOCK JURORS AND JURIES

By

Robert J. MacCoun

A DISSERTATION

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department Of Psychology

1984

ABSTRACT

MODELING THE IMPACT OF EXTRALEGAL BIAS AND DEFINED

STANDARDS OF PROOF ON THE DECISIONS OF MOCK JURORS AND JURIES

By

Robert J. MacCoun

Research on the psychology of the courtroom has documented
extralegal bias -- bias due to inadmissable or non-evidentiary factors
-- in the verdicts and related judgments of mock jurors. This
dissertation describes a criterion-setting model proposing that
extralegal factors influence the juror's standard of proof, the
threshold of evidence that must be crossed to render a guilty verdict.
The relationship between bias and this criterion is hypothesized to be
mediated by the perceived costs of convicting an innocent defendant or
acquitting a guilty one.

A 2 (Defined Standard of Proof) x 2 (Victim Attractiveness) x 2
(Defendant Attractiveness) x 2 (Subject Sex) factorial experiment was
conducted. Three hundred and twenty one subjects participated in
a simulated auto theft trial. Subjects were shown photographs that
varied in physical attractiveness and allegedly portrayed the victim
and the defendant. Subjects received either "beyond a reasonable
doubt" or "mere preponderance of evidence" standard of proof
instructions.

As predicted, the instructional manipulation resulted in a higher
conviction rate for the preponderance of evidence standard than for
the reasonable doubt standard. Although the case was close and the
attractiveness manipulations were strong, this study was not

able to detect attractiveness effects on either pre-deliberation verdicts or recommended sentences. This failure to replicate previous research might have resulted from the addition of auditory trial information.

Each of several criterion estimates was more accurate at predicting verdicts than expected by chance; however, the decision theory and rank-order procedures were each significantly more accurate than three self-reported probability formats, which may have been inflated by social desirability.

Despite the absence of bias prior to deliberation, groups were significantly less likely to convict the defendant when he was attractive. This pattern is in clear contradiction to Kaplan and Miller's (1978) hypothesis that deliberation should attenuate extralegal biases. Criterion modeling revealed more stringent criterion estimates in the attractive defendant condition. There also an unexpected opposite trend for the perceived weight of evidence.

Social Decision Scheme analyses demonstrated an asymmetry effect for group verdicts; however, the hypothesis that this asymmetry results from the reasonable doubt standard was not supported.

Dedicated

to the Memory of

Barbara A. MacCoun

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In the psychological study of the legal system, it is the squeaky wheels that receive the grease. As in other areas of psychology, legal psychologists tend to focus on pathology. Although this focus on flaws and problems at times may strike the layman as a morose outlook on the world, for the psychologist, it is often the simplest way of gleaning insights into the way things work when they work well. For instance, our system of common law is a system of fact-finding and fact-weighing. Its personification, Themis, balances facts in her scales and shields her eyes from all appearances which threaten to seduce and mislead. But for the psychologist, the obvious task is to stand at her feet and try to catch her peeking. Thus, up to this point in its relatively brief history, the psychology of the law has been predominantly the study of extralegal bias, whether in eyewitness testimony, pre-trial publicity, parole decisions, jury composition, or the impact of physical and personal characteristics of the plaintiff and the defendant on the administration of justice.

Characteristics of the actors in the system may, in some instances, be legally relevant (e.g., credibility), but psychologists have tended to focus on extralegal characteristics, i.e., characteristics that should have no legal bearing on the decision to convict or acquit the defendant. For example, Sigall and Ostrove (1975) have demonstrated that jurors are influenced by the physical attractiveness of the defendant. Kerr (1978a) found that mock jurors were more likely to vote for conviction when the victim of a crime was

1

both "beautiful and blameless" (i.e., when she took precautions to prevent the crime). Other variables whose potential effects have been explored include the race, religion, occupation, and physical stigmata of defendants and victims. (For a recent review of this literature, see Dane & Wrightsman, 1982.)

The tacit assumption behind the extralegal status of these characteristics seems to be that they should have no objective logical bearing on the weight of evidence against the defendant. For legal theorists, the question is: Does bias miscalibrate the balance or come to rest on its scales? Some scholars (e.g., Kaplan & Miller, 1978; Shaffer, Case, & Brannen, 1979) argue that bias is weighed along with the evidence -- a disturbing prospect but for the hope that the evidence can come to weigh increasingly more, relative to the biasing information, as the judgment process proceeds. Others (e.g., Kerr et al, 1984; Thomas & Hogue, 1976) have argued that when bias influences verdicts, it often does so through its impact on the judge or juror's standard of proof, the criteria for the amount of evidence necessary to conclude guilt beyond a reasonable doubt. This dissertation examined the potential extralegal impact of victim and defendant attractiveness in a mock criminal trial, and in addition, it explored the role of two legal procedures -- the judge's charge to the jury and the jury deliberation -- in shaping jurors' judgments and possibly moderating extralegal bias.

Juror Decision-Making and the Hypothesis-Testing Metaphor

The Thomas and Hogue model. In 1976, Thomas and Hogue presented a formal mathematical model of juror decision-making that is roughly analogous to formal models in signal detection theory. But since the

2

"true" state can never be known in legal hypothesis-testing, the Thomas and Hogue model includes only one distribution.

Thomas and Hogue have postulated two relevant parameters: the perceived weight of evidence against a defendant, and a judgmental criterion for "reasonable doubt." The perceived weight of evidence is conceived of as a random variable, X, with probability density function $f(x)$, and expected value, m. The decision criterion, c, divides $f(x)$ into regions "for" and "against" the defendant. Thus, the ith juror will compare his/her estimate of the defendant's guilt, $X(i)$ against the criterion, and will convict if $X(i) > c$, or acquit if $X(i) < c$.

In order to estimate m and c, Thomas and Hogue make the assumption that a juror's confidence in his/her verdict is a monotonically increasing function, g, of $|X - c|$, the discrepancy between the perceived weight of evidence and the decision criterion. This allows them to estimate m and c by collecting jurors' verdicts and ratings of confidence-in-verdict. As a matter of mathematical convenience, Thomas and Hogue further assume that $f(x)$ is characterized by an exponential and asymmetric distribution. Through a rather complex bootstrapping operation which is beyond the scope of this paper (and, at the moment, beyond the mathematical prowess of its author), they compare and evaluate three such distributions (exponential, generalized gamma, and generalized Laplace) and demonstrate that these assumptions are reasonably valid.

Using Thomas and Hogue's model, Kerr (1978a) has demonstrated that the impact of the attractiveness and precautiousness of the victim on mock jurors' verdicts was mediated by shifts in their reasonable doubt

3

criterion. When the victim was "beautiful and blameless," jurors required less evidence to convict the defendant than when she was not. In a second application, Kerr (1978b) demonstrated that the conviction rate for mock jurors was inversely related to the severity of the prescribed penalty; again, the Thomas and Hogue model indicated that this relationship was mediated by jurors' requirements of proof. Kerr, Bull, MacCoun, and Rathborn (1984) found that a victim's attractiveness, precautiousness, and degree of facial disfigurement influenced mock jurors' verdict decisions, and that this effect appeared to be mediated by the reasonable doubt criterion.

In order to get an intuitive grasp of the Thomas and Hogue model, it is useful to consider a metaphor that is familiar to most psychologists. Feinberg (1971) has pointed out that the juror's task is very similar to that of the inferential statistician. Both attempt to infer "truth" based upon the available evidence. The common law notion that the defendant is "innocent until proven guilty" provides the juror with a "null hypothesis," and the "reasonable doubt" criterion provides the juror with an "alpha level." Thus, Feinberg points out that the juror faces two possible errors: the "Type I error" of convicting an innocent defendant, and the "Type II error" of acquitting a guilty defendant. Although Feinberg originally conceived of this metaphor as a tool for teaching statistics to college undergraduates, many psychologists have found it to be a useful tool for understanding and modeling the juror's task.

Following Feinberg's metaphor, Kerr, et al.(1984) argued for an additional mediational link in the juror's decision process. We suggested that extralegal victim and defendant characteristics may influence the reasonable doubt criterion by affecting the perceived

4

costs of acquitting a guilty person or convicting an innocent person, respectively. This conceptual framework is illustrated in Figure 1. Factors that lead jurors to sympathize with a defendant might heighten their concern over avoiding a false conviction. As an extreme example, consider a juror who is a personal friend of the defendant. We would argue that this juror would require a great deal more evidence to vote to convict than would a juror who was a complete stranger. However, we would not expect these jurors to differ in their reaction to seeing the defendant set free, if the weight of evidence clearly suggests that he is guilty.

Factors which lead jurors to sympathize with the victim of a crime, on the other hand, should lead to an increase in their desire to avoid acquitting a guilty person. Consider a second example (again, a rather extreme one) in which a juror is a friend of the victim of a rape. We would predict that this juror will be much more concerned about the possibility of acquitting the defendant if he is guilty, especially since he may retaliate against the victim, than would a juror who was a complete stranger. However, we would not expect these jurors to differ in their desire to avoid convicting the defendant if he is clearly innocent.

We are suggesting that these perceived costs may be reflected in the level at which jurors set their reasonable doubt criterion. In extreme cases, it is possible that jurors might either lower their criterion so low as to convict the defendant no matter what the evidence, or raise it so high as to refuse to convict an obviously guilty defendant. However, the likelihood of such extreme cases may be minimized by the voir dire procedure.

Figure 1
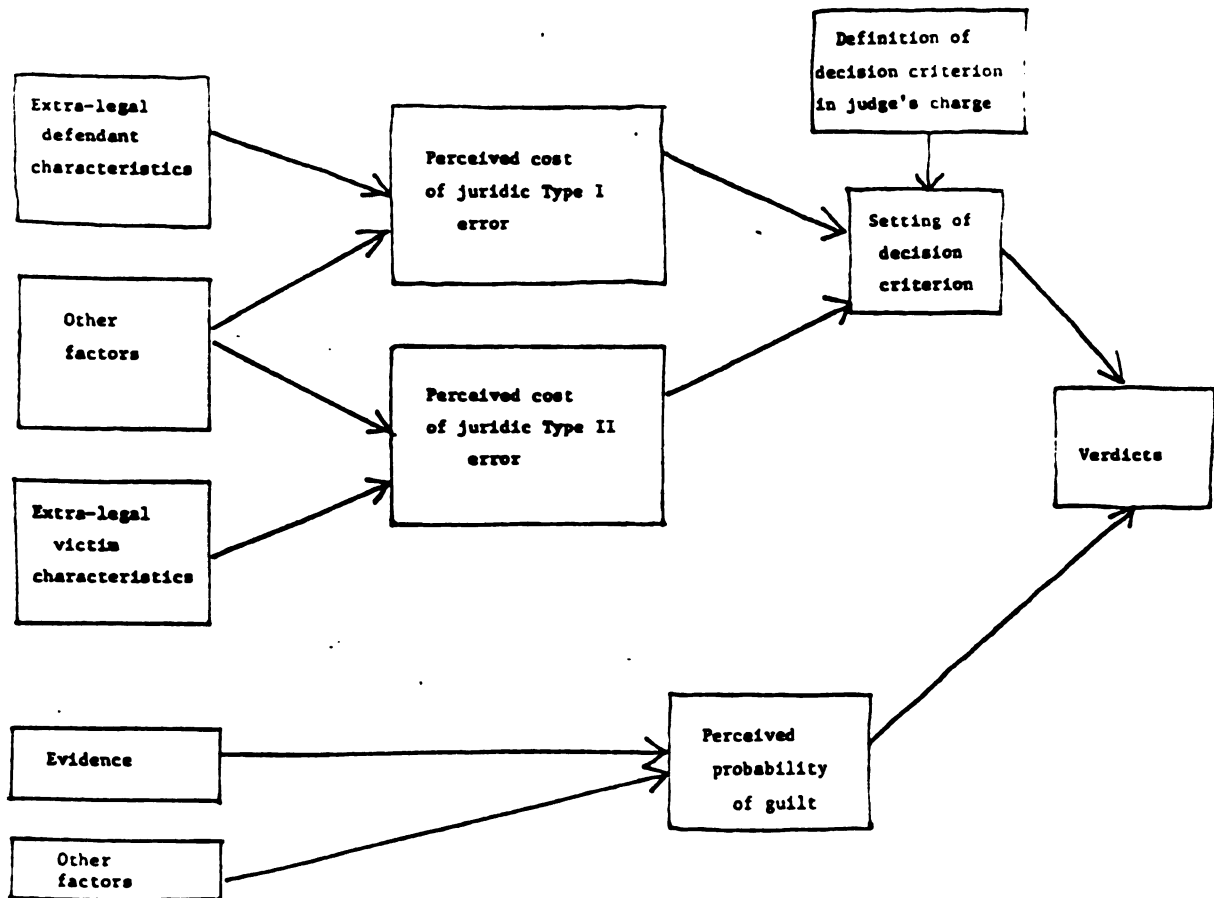
It seems more likely that extralegal victim and defendant
characteristics will manifest themselves more subtly, and will only
effect verdicts when the evidence is rather equivocal, i.e., near the
range of most jurors' reasonable doubt thresholds. Informally, we can
suggest a number of extralegal factors that might plausibly influence
the perceived costs of these juridic errors in actual practice. For
example:


```
PERCEIVED            ──────*Attraction to/sympathy for defendant
COST OF          ◄──────────*Penalty severity
TYPE I ERROR                 *Demand for improvement in police
                              inquiry or conduct
                            *Belief in efficacy/morality
                             of penal system


PERCEIVED            ──────*Attraction to victim
COST OF          ◄──────────*Belief in deterrent effect
TYPE II ERROR               *Desire to avenge victim
                            *Desire to punish defendant
```
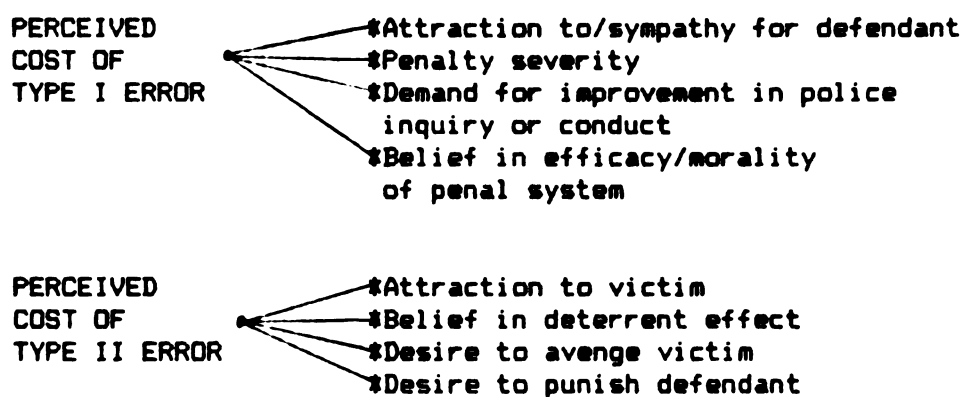

We would not expect all extralegal factors to influence verdicts
through the mediation of the reasonable doubt criterion, however. For
example, jurors' verdicts could be influenced by evidence (e.g., prior
criminal record) that has been ruled inadmissable during the trial
(cf. Doob & Kirshenbaum, 1972; Hans & Doob, 1976; Sue et al, 1973,
1974) or has been publicized prior to the trial and then excluded from
the trial (cf. Kerr & MacCoun, 1983). Such factors could plausibly
effect the perceived weight of evidence without influencing the
decision criterion.

Figure 1 also suggests that in addition to these juridic costs,
the judge's reasonable doubt instructions to the jury could also serve
as an input in establishing the stringency of the decision criterion.

7

By instructing the jurors to set a very stringent criterion level, the judge can create the same result that a high concern over Type I errors would have. And indeed, there is a good reason for doing so. As Loftus (1983) has pointed out, the Type I error may indeed be more costly, for it is easy to overlook the fact that when we convict an innocent defendant, we also neglect to convict the true culprit. Ironically, Champagne and Nagel (1982) have reviewed a number of political reasons why judges may tacitly prefer a policy that reduces the decision criterion, despite the risk of Type I errors that result in the conviction of an innocent defendant while the true culprit remains at large.

It is important to consider an additional effect that judicial instructions may have, however. In addition to simply raising or lowering the criterion level, these instructions may also change jurors' perceived costs of Type I and Type II errors, or reduce the weights that jurors place on these subjective factors. Thus, the instructions could actually reduce the impact of extralegal victim and defendant characteristics, regardless of the direction of their effects.

An Overview

The present chapter reviews research relevant to the conceptual model presented in Figure 1, including models of the evaluation of evidence, attempts to quantify the reasonable doubt criterion, manipulations of the judicial definition of reasonable doubt, and studies of factors influencing jurors' ability and motivation to comply with judicial instructions. Then, subsequent chapters will describe a 2 (Victim attractiveness) X 2 (Defendant

8

attractiveness) X 2 (Standard of Proof instructions) factorial experiment that was designed to provide a direct test of various components of the model. Although any number of extralegal victim and defendant characteristics might be useful for validating the model, physical attractiveness has the benefit of being unambiguously extralegal in an auto theft case such as that used here.

In addition to assessing individual verdicts and related judgments, the present study also examined the verdicts of deliberating jurors. Assessments of group verdicts must employ the group as unit of analysis; unfortunately, practical constraints prohibit the use of traditional twelve-person juries if the analyses are to have adequate statistical power. Therefore, in the present study, subjects deliberated in groups of two to four after completing the individual questionnaires. The use of deliberating groups allowed tests of several hypotheses regarding (a) the impact of deliberation on extralegal biases (cf. Kaplan & Miller, 1978), and (b) the asymmetry typically found in social decision scheme matrices (cf. Stasser, Kerr, & Bray, 1982). These hypotheses are described in greater detail below.

The Extralegal Effects of Victim and Defendant Attractiveness

Preliminary experimental evidence of the long-standing hunch that good-looking defendants can "get off easy" came from a study reported by Efran (1974). Efran provided students with a photograph of an attractive defendant, an unattractive defendant, or no photograph at all, and a fact sheet describing an incident in which a student was allegedly caught in the act of cheating. He found that the attractive defendant was less likely to be found guilty, and received a lighter

9

punishment than the less attractive defendant. Unfortunately, Efran counfounded the sex of the subject with the sex of the defendant by providing only male photos for females, and vice-versa. This confounding is unfortunate because post-hoc contrasts suggest that the effects for guilt and punishment were only significant for males judging females, a pattern which is therefore difficult to interpret.

Kaplan and Kemmerick (1974) used trait adjectives to manipulate the social attractiveness of defendants. Defendant characterization and the amount of evidence were both varied in a within-subject design employing a series of traffic felony trials. In addition, one third of the subjects were told that the nonevidentiary defendant characterizations might be useful for their judgments, one third were told that such information was often misleading and inaccurate, and the remaining third were given no special instructions. Kaplan and Kemmerick report that both the evidentiary and nonevidentiary factors were integrated in an additive fashion consistent with the predictions of a weighted-average model described in a later section of this chapter. The instructional manipulation had no effect, however.

A number of studies of defendant attractiveness (e.g., Izzett & Fishman, 1976; Izzett & Leginski, 1974; Michelini & Snodgrass, 1980; Sigall & Ostrove, 1975) have examined possible moderating variables. Sigall and Ostrove (1975) examined the possible moderating influence of type of crime. Subjects read a brief transcript of a trial in which a woman was either accused of (a) burglarizing an apartment for $2,200 in cash and merchandise, or (b) swindling a middle-aged bachelor into investing $2,200 in a non-existant corporation. Subjects received either a photograph of an attractive defendant, an unattractive defendant, or no photograph at all. The investigators

solicited an attractiveness manipulation check and subjects'
recommended prison sentences, but regretably neglected or chose not to
obtain guilt judgments. Thus, jurors predisposed toward acquittal had
a limited range of response options for reacting to the trial
stimuli. Since the trials were constructed to imply the defendant's
guilt, however, this might not have been a serious problem.

An Attractiveness by Offense interaction indicates that subjects
were significantly more lenient with the attractive defendant if the
crime was burglary, but more lenient with the unattractive defendant
if the crime was a swindle. Simple effects tests indicated that the
latter comparison was not significant, however. Comparisons to the
control condition suggested that the defendant received almost
identical treatment in both the unattractive and no photograph
conditions; apparently, unattractive defendants did not receive
discriminatory treatment in either case. Sigall and Ostrove argue
that the swindle case was a crime for which attractive defendants are
more likely to be successful and more likely to pursue in the future.
Conversely, for a more conventional type of crime, attractive
defendants receive the benefit of the doubt because they presumably
have socially desirable traits which would promote rehabilitation and
successful adjustment to the community.

Sigall and Ostrove interpret their pattern of results as
supporting a cognitive rather than a reinforcement-affect
interpretation of attractiveness-leniency effects. In this study, it
is difficult to determine whether attractiveness influenced subjects'
estimates of the probability of guilt or their judgments of the
expediency of rehabilitation. This difficulty is compounded by (a)

the reliance on sentencing guidelines as a primary dependent measure, and (b) the apparent confounding of specific photographs with levels of attractiveness, so that a specific facial cue have triggered inferences of intelligence, expressivity, or some other trait assumed to be relevant to the burglary/swindle distinction.

Michelini and Snodgrass (1980) follow a similar line of reasoning. Subjects in their study received descriptions of defendants which were either positive or negative and either relevant or irrelevant for a traffic felony. Michelini and Snodgrass found that attractive traits only reduced perceptions of guilt when those traits had relevance to the crime (e.g., "careful and deliberate").

However, these results are also problematic. Since different positive or negative traits were used depending on whether they were relevant or irrelevant, these two factors were not truly crossed. The manipulation check for relevancy solicited from a pilot group of subjects indicated a reliable trait attractiveness by relevancy interaction. Decomposition revealed a simple main effect for attractiveness for relevant traits, in which attractively described people were expected to be less likely to act in the described criminal manner than unattractively described people. No main or simple main effects for relevancy on its own manipulation check are reported, and the subjects in the main study did not complete manipulation checks.

Less experimental attention has been given to the effects of victim attractiveness. While Thornton (1977) found no effects for guilt ratings, the experiment involved a rape trial, in which victim attractiveness may have had implications for probability of guilt which might counteract any tendency to help an attractive victim find

12

justice. As described above, Kerr (1978) found that conviction was more likely in an auto theft case when the victim was attractive, but only if she took necessary precautions to avoid the crime.[1] Kerr, Bull, MacCoun, and Rathborn (1984) found that a complex interaction of victim attractiveness, precautiousness, and facial disfigurement influenced guilt ratings, but found no main effect for attractiveness.

To summarize, although numerous studies report defendant attractiveness effects, these studies are plagued by a myriad of methodological flaws, and they are limited in some cases to sentencing effects rather than guilt effects. Moreover, several of these studies manipulated social, rather than physical, attractiveness. Relatively few studies have examined victim attractiveness, and those that have do not present a simple pattern. Therefore, a first objective of the present study was to examine whether victim and defendant attractiveness have reliable effects upon judgments of guilt.

Modeling the Evaluation of Evidence

Thomas and Hogue (1976) do not articulate the cognitive processes involved in evaluating the evidence presented in a trial in order to establish a perceived weight of evidence. They do rationalize the use of an exponential pdf $f(x)$ by assuming a Poisson process in which apparent weight of evidence increases by amount $k\Delta$ every time interval of length Δ until some critical evidentiary datum appears at the interval from t to $t + \Delta$, at which point apparent weight "freezes" at $X = kt$. Thomas and Hogue seem to suggest that this assumption was created as a mathematical convenience rather than a psychological postulate. In their review of juror decision-making models, Pennington and Hastie (1981; cf. Penrod & Hastie, 1979) review several

13

other, more specific, cognitive models of juror decision-making. I will review two such models briefly.

Information Integration Model. Based on the work of Norman Anderson (e.g., 1981) the Information Integration model suggests that jurors combine their initial estimate of guilt with information presented during the trial in a process of valuation, the assignment of scale values (s_i) to each piece of information, and integration, in which each piece of information is given a weight (w_i) and averaged. For example:

$$J = \frac{w_0 s_0 + \Sigma s_i w_i}{w_0 + \Sigma w_i} \quad\quad [1]$$

where J is the subjective likelihood of guilt. Ostrom, Werner and Saks (1978) have used this approach to analyze mock jurors' presumption of innocence. They distinguish between four possible juror strategies of "fair mindedness": (1) the juror can set $s_0 = .50$ deciding that guilt or innocence are equally likely; (2) the juror can attempt to be objective, and since more persons brought to trial are found guilty than not-guilty, the juror can set $s_0 > .50$ (see the Bayesian model, described below, for an analysis of this type of reasoning); (3) the juror can actually "presume innocence", i.e., set $s_0 = 0$; or (4) the juror can decide to completely ignore his or her predispositions, whatever they may be, by setting $w_0 = 0$. Ostrom et al. report that their mock jurors did apparently presume innocence (strategy 3), and that s_0 was averaged with the trial evidence to produce a judgment of guilt. They also classified subjects as either

14

pro- or anti-defendant, and found that while anti-defendant subjects actually set a lower level of s_0, they were also quicker to abandon the presumption of innocence in the face of evidence than were pro-defendant subjects. The latter result may indicate that the anti-defendant subjects had a lower reasonable doubt criterion; however, a drawback of the information integration model is that it does not account for the reasonable doubt criterion (cf. Pennington & Hastie, 1981), making it an incomplete portrayal of the juror decision process.

Pennington and Hastie (1981) also review a sequential weighing model that is an earlier precursor to the information integration model. Its primary distinction is that it assumes that the averaging process takes place sequentially, as each new item is encountered, rather than at the completion of the trial. This assumption makes the sequential weighing model more consistent with the Poisson process assumed by the Thomas and Hogue model.

Martin Kaplan (e.g., Kaplan & Miller, 1978) has been one of the foremost proponents of the information integration model. He has provided a conceptualization of the deliberation process which yields strong predictions regarding the impact of extralegal factors on verdicts. Kaplan's work is detailed in a later section of this chapter.

Bayesian Model. The Bayesian model is a normative model suggesting the correct approach to integrating evidentiary information consistent with probability theory. As such, it may not be a good model of how jurors actually do reach decisions. Such models are often used both as theories of decision-making (when they fit the subjects' data) and as tools for discovering cognitive biases (when they don't).

One such model (Marshall & Wise, 1975) is:

$$R_n = P(G|E_n)/P(NG|E_n) \qquad [2]$$

where R_n is the posterior odds for guilt, the ratio of the probability of guilt given all the evidence to the probability of not guilty given all the evidence; this model is algebraically equivalent to the probability of guilt prior to the evidence, R_0, multiplied by the product of the likelihood ratios for each item of evidence, which are measures of the diagnosticity of each item of information for assessing the probability of guilt. As with other applications of Bayesian analyses (cf. Nisbett & Ross, 1980), this model has not described the decisions of mock jurors very accurately (Pennington & Hastie, 1981).

Modeling the Decision Criterion

The American judicial system has adopted the common-law tradition of protecting the defendant from false conviction by placing the "burden of proof" in a criminal trial upon the prosecution. The defendant is to be "presumed innocent until proven guilty." During the "fact-finding" process of the trial, the prosecution presents evidence against the defendant in an attempt to build a case establishing guilt. At the conclusion of the trial, jurors (or, in a bench trial, the judge) must review the evidence and decide to convict the defendant if, and only if, the evidence indicates "beyond a reasonable doubt" that the defendant committed the crime.

Unfortunately, the reasonable doubt criterion, although noble in spirit, is extremely vague and difficult to define unambiguously in

16

practice. A variety of "stock" definitions have been created in various American court systems, and in some courts the judge is given discretion to define the criterion as he or she sees fit in a specific trial. As Simon (1970) has suggested, judges often attempt to define the phrase "beyond a reasonable doubt" by providing jurors with paraphrases or apparently synonymous terms. For example:

> Reasonable doubt is one a reasonable person has after carefully weighing all the testimony and is one a reasonable person would act or decline to act upon. It is not a capricious doubt or a fanciful doubt or a doubt arising in anyone's mind because of any sympathy for the defendant. It is in essence what the words obviously mean - a reasonable doubt. A reasonable doubt may arise not only from the evidence produced but also from a lack of evidence.

Numerous researchers have attempted to translate the reasonable doubt criterion into a more concrete, quantifiable definition. One approach, the Thomas and Hogue (1976) model, has already been described above. Several other approaches are reviewed below.

The Blackstone approach. Following Blackstone's assertion that "it is better to let ten guilty men go free than to allow one innocent man to be convicted" (cited in Kaplan, 1982), several authors have suggested that the reasonable doubt criterion can be expressed as such a ratio. For example, Grofman (1977) argues that in order to minimize the expected disappointment in a verdict, jurors should rationally apply the formula

$$p* = \frac{r}{r + 1} \qquad \qquad [3]$$

where p* is the threshold probability above which the juror is able to convict beyond a reasonable doubt, and r is the number of guilty defendants the juror is willing to set free in order to avoid convicting one innocent defendant. Kaplan (1982) points out that Blackstone's assertion, often considered representative of the viewpoint of the American judicial system, therefore sets the criterion at .91. Jurors can therefore compare their subjective probability of guilt estimate against this criterion and convict if and only if p(G) exceeds the criterion (cf. Cullison, 1977).

The Self-Report approach. Simon and Mahan (1971) operationalized the decision criterion as the minimum probability of guilt required for a given judge or juror to vote guilty. Respondents were asked the following question:

What would the likelihood or probability have to be that a defendant committed the act for you to decide that he is guilty? (FILL IN THE BLANK)

I would have to believe that it was a _____ out of ten chance that the defendant committed the act.

Simon and Mahan solicited self-reported criteria from judges, members of the jury pool, and college students and found mean probabilities of .89, .79, and .89, respectively. The modal criterion reported by each group was 1.0, a requirement of absolute certainty. The fact that 31% of the judges required absolute certainty and 69% did not may

18

demonstrate the ambiguity of the "beyond a reasonable doubt" concept;
alternatively, it may indicate that respondents had a difficult time
using Simon and Mahan's response format.

Iversen (1971) has criticized Simon and Mahan's operationalization
of the concept of probability, arguing that it is meaningless to use
probability in the sense of "relative frequency," since the defendant
can not be tried ten times for the crime (nor, for that matter, can we
try ten different individuals with identical evidence in the same
trial). Instead, Iversen advocates the use of an "uncertainty"
conceptualization of probability, in which numbers between zero and
one signify the degree to which the juror is certain as to the
defendant's guilt. (See Kerr, et al, 1976, and Dane, 1979, for
applications using the zero-to-one scale).

The Rank-Order approach. Simon (1967) had half of her mock
jurors indicate their verdict after reading a trial transcript; the
other half were instructed to indicate the probability that the
defendant committed the act for which he was charged, using a 21-point
scale ranging from "0 out of 10 chance" to "10 out of 10 chance."
Simon then rank-ordered the probabilities from highest to lowest.
Assuming that the subjects in both groups were randomly distributed,
Simon obtained an estimate of the reasonable doubt criterion for the
sample by counting down the probabilities, until the number of guilty
votes in the other group was reached. Using this technique, Simon
found estimates of .70 to .74. Unfortunately, as Dane (1979) has
pointed out, the accuracy of this technique could not be assessed since
Simon did not obtain both measures from each subject.

The Statistical Decision Theory approach. This approach (e.g., Fried, Kaplan, & Klein, 1975; J. Kaplan, 1968; Nagel, 1979, 1982) is more theoretical than the previous, methodological approaches to modeling the decision criterion. Furthermore, it is similar to the conceptual model of Kerr, et al. (1984), because it includes juror's perceived costs of Type I and Type II errors as components of the decision criterion. In addition, it also considers the perceived utilites of correct verdicts. One such model was offered by Fried, Kaplan, and Klein (1975), who consider the following matrix of subjective expected utilities (U's):

	State of the World	
	Guilty	Innocent
Convict	U_{CG}	U_{CI}
Decision		
Acquit	U_{AG}	U_{AI}

Note that UCI and UAG correspond to what Feinberg (1971) and Kerr et al. (1984) refer to as Type I and Type II errors, and are conceptualized as "disutilities" with a value less than zero. Fried et al. suggest that the expected utility (EU) of convicting the defendant is

$$EU(C) = p U_{CG} + (1-p) U_{CI} \qquad [4]$$

where p is the juror's subjective probability of guilt estimate. Similarly, the expected value of acquitting the defendant is

$$EU(A) = p\,U_{AG} + (1-p)\,U_{AI} \qquad [5]$$

Fried et al. argue that a juror should convict if and only if

$$EU(C) > EU(A) \qquad [6]$$

or

$$p\,U_{CG} + (1-p)\,U_{CI} > p\,U_{AG} + (1-p)\,U_{AI} \qquad [7]$$

Algebraically, Fried et al. then proceed to derive the juror's decision rule given the above assumptions:

$$p(U_{CG} - U_{AG}) + (1-p)(U_{CI} - U_{AI}) > 0 \qquad [8]$$

$$p(U_{CG} - U_{AG}) + (U_{CI} - U_{AI}) + p(U_{AI} - U_{CI}) > 0 \qquad [9]$$

$$p(U_{CG} - U_{AG} + U_{AI} - U_{CI}) > U_{AI} - U_{CI} \qquad [10]$$

$$p > \frac{U_{AI} - U_{CI}}{U_{CG} - U_{AG} + U_{AI} - U_{CI}} \qquad [11]$$

Thus, the right half of equation [11] represents the decision criterion, which Fried et al. denote as p* in equation [12]:

$$p* = \frac{U_{AI} - U_{CI}}{U_{CG} - U_{AG} + U_{AI} - U_{CI}} \qquad [12]$$

Fried et al (1975) provide two hypothetical examples of how this

formula might model a juror's reasonable doubt criterion. First, they

consider a juror who believes that the penalties for a crime like

possession of marijuana are overly severe (cf. Kerr, 1978b). Such a

juror might have the following utility matrix:

```
                        State of the World
                        --------------------
                        Guilty      Innocent
                        --------------------
              Convict     10          -5000
    Decision
              Acquit       0           100
```

Applying equation [12] for this juror, we find an extremely stringent

decision criterion:

$$
p* = \frac{100 - (-5000)}{10 - 0 + 100 - (-5000)} = \frac{5100}{5110} = .998
$$

Next, Fried et al. consider a juror in a rape trial in a community in

which there has been a recent wave of rapes. Such a juror might have

the following utility matrix:

```
                        State of the World
                        --------------------
                        Guilty      Innocent
                        --------------------
              Convict     100         -1000
    Decision
              Acquit     -200          300
```

Applying equation [10], Fried, et al. estimate this juror's criterion

as p* = .82, a much more lax standard.

Note that Fried, et al. (1975) do not provide an explicit link

between the judge's defined standard of proof and the juror's

functional criterion. One possibility is that the judge's

instructions have an indirect influence upon the functional criterion by influencing jurors' utility estimates for the four possible trial outcomes. This issue is addressed in more detail in a later section of this chapter.

The comparative accuracy of the approaches. Except for the Thomas and Hogue (1976) model, all of these approaches to quantifying the reasonable doubt criterion place it on a zero-to-unity metric. This is convenient because it allows the researcher to compare the criterion to each mock juror's subjective probability of guilt estimate in order to create an "expected verdict." This expected verdict may then be compared to the mock juror's actual verdict in order to assess the accuracy of the operationalization of "reasonable doubt"; i.e., the method will either "hit" or "miss."

Dane (1979) utilized this technique for comparing several alternative estimates of his mock jurors' reasonable doubt criteria following a trial simulation. The mean criterion estimates for the Statistical Decision Theory (SDT), self-report, and rank-order approaches were .52, .66, and .73, respectively. Dane found that the rank order estimates were approximately 88% accurate, the SDT estimates were approximately 82% accurate, and the self-report estimates were approximately 77% accurate. All the estimates were significantly more accurate than expected by chance. It is not surprising that the rank-order estimates achieved such a high level of accuracy, since the rank-order approach and the hit rate procedure are both premised upon the positive monotonic relationship between the conviction rate and $p(G)$. The rank-order procedure has an advantage over the SDT procedure in that it may be more generally

applicable; Dane reports that for 11 of his 168 mock jurors, the SDT estimate fell outside the zero-to-unity range, as a result of jurors assigning positive utilities to CI or AG, negative utilities to CG or AI, or both. To the extent that its assumptions are valid, the Thomas and Hogue model has the advantage that it does not require the use of subjective probability and expected utility estimates, which are suspected to be very difficult for mock jurors to make (cf. Kerr, et al., 1984).

Dane also examined the correlations between the criterion estimates and mock jurors' confidence-in-verdict ratings, and he found mixed support for Thomas and Hogue's assumption of a positive, monotonic relationship. While direct support would be encouraging, mixed or non-support can only be inconclusive, since we do not know whether the Thomas and Hogue model is invalid, or whether the alternative approach used to derive the criterion is invalid. Without independent evidence for Thomas and Hogue's assumed $g(|X - c|)$ function, such bootstrapping remains problematic.

The present study solicited subjects' verdicts, confidence-in-verdicts, subjective probability of guilt estimates, self-reported criterion estimates, and perceived outcome utilities and costs. This allowed a comparison of the Blackstone, SDT, Thomas and Hogue, Rank-Order, and Self-Report estimates of the decision criterion for accuracy, stringency, and validation of the model presented in Figure 1. Individual estimates of the criterion, as provided by the Blackstone, SDT and Self-Report approaches, permitted standard parametric test of the hypotheses. On the other hand, research described below suggests some potential methodological artifacts that can result from reliance on self-report estimates. Therefore, Thomas

24

and Hogue and SDT estimates are useful as an independent check
on such problems. Finally, it seems safe to anticipate that subjects
may have some difficulty providing reliable subjective probability and
expected utility estimates. Therefore, it seemed prudent to (a)
collect several different estimates and seek some convergence, and (b)
use large cell sample sizes, in order to increase the sensitivity of
the analyses.

Experimental Research on Judicial Instructions

Kerr, Atkin, Stasser, Meek, Holt and Davis (1976) manipulated the
judge's charge to the jury in a trial simulation. They provided
subjects with either no definition of reasonable doubt, a lax
definition ("...a reasonable doubt must be a substantial one, ... one
for which reasons can be given ... you need not be absolutely sure
that the defendant is guilty to find him guilty"), or a stringent
definition ("...if you feel that the facts of this case are compatible
with any other theory of this case besides the one in which the
defendant is guilty, then you have a reasonable doubt..."). Kerr, et
al. demonstrated that the variations in criterion definition had a
significant impact on both pre- and post-deliberation verdicts, with
the largest proportion of guilty verdicts obtained in the lax
condition (60% and 62%, pre- and post-deliberation), followed by the
no definition condition (51% and 57%), and finally, the fewest
convictions in the stringent condition (46% and 35%). Using the self-
report approach, Kerr, et al. found mean criterion estimates of .87, .82,
and .82 for the stringent, lax, and no definition conditions. The mean
for the stringent definition was significantly greater than for the
other two definitions (p < .005). Furthermore, Thomas and Hogue

(1976) estimated c for each condition in the Kerr, et al. (1976)

study, and found a similar pattern of decision criteria.

Nagel (1979; see also Nagel, Lamm, & Neef, 1981; Nagel & Neef,

1979) also varied the content of the judge's definition of the

decision criterion in a simulated rape trial. He compared a no-

definition control condition with a "beyond a reasonable doubt"

definition, a ".90 probability" definition, and Blackstone's "10:1

Tradeoff" definition. Nagel estimated each subject's criterion using a

variation of Fried et al.'s SDT approach. First, subjects were asked

which of the four possible trial outcomes (i.e., AI, AG, CI, CG) they

considered to be desirable and which they considered to be

undesirable. Nagel reports that most subjects considered both AG and

CI undesirable, and felt that CI (the Type I juridic error) is more

undesirable than AG. Subjects were then asked to place the most

undesirable outcome at -100 on a 0 to -100 scale, and then to place

the second most undesirable outcome on the scale between -100 and 0,

at a value Nagel denotes as "X." By making the simplifying assumptions

that $|UAI| = |UCI|$ and that $|UAG| = |UCG|$, Nagel then calculated the

criterion by using the following formula:^{2}

$$p* = \frac{100}{(100 - X)} \qquad [13]$$

For example, Nagel (1979) argues that Blackstone would presumably

consider AG one tenth as bad as CI, so that X = -10, yielding .91, as

in Grofman's formula [3].

Using this approach, Nagel (1979) reports the following mean

estimates:

	Males	Females
No Instructions	.70	.50
"Beyond a Reasonable Doubt"	.75	.60
.90 Probability	.80	.75
10:1 Tradeoff	.90	.90

Note that females were generally more lax than males. Although the sex differences were not statistically significant, they suggest that females appear to be more predisposed to convict the defendant in a rape trial, as we would expect for females based upon the use of expected cost estimates. It would be interesting to assess whether or not the self report, rank order, and Thomas and Hogue approaches would also reflect such differences. If so, it might suggest that the SDT approach is simply an explicit model of what the other approaches model implicitly.

Curiously, the .90 probability and 10:1 tradeoff instructions define almost identical ideal criteria (.90 and .91, respectively), and yet differ considerably in the estimates they yield. Not only were the estimates in the 10:1 condition closer to the defined ideal, they also did not reflect the sex differences apparent in the other conditions. This pattern of mean criterion estimates may result from differences in the comprehensibility of the different judicial instructions, or differences in mock jurors' motivation to comply with the instructions. Each of these possibilities will be addressed below.

Comprehensibility. Research by Charrow and Charrow (1979) and by Sales, Elwork, and Alfini (1977) has demonstrated that most jurors

only understand a small percentage (under 50 %) of the instructions
that are read to them. Typical judicial instructions are often legally
precise but semantically vague, archaic, or redundant, and, as such, often
can be reworded so as to greatly improve their comprehensibility.
Consider, for example, instruction 3.71 from the Book of Approved Jury
Instructions:

> If you should find that John Smith, who, at the time of the
> accident in question, was driving the vehicle in which plaintiff
> was riding, was negligent and that his negligence contributed as
> a proximate cause of plaintiff's injury, then you must determine
> whether said driver was then the agent of the plaintiff and
> acting within the scope of his employment.

> If the driver was plaintiff's agent and acting within the scope
> of his employment, his negligence, if any, must be imputed to the
> plaintiff, with the same effect as if the plaintiff himself were
> contributorily negligent.

> But if said driver was not then the agent of plaintiff or was not
> acting within the scope of his employment, his negligence, if
> any, may not be imputed to the plaintiff.

Compare those instructions to the modified version constructed by
Charrow and Charrow (1979):

> As you recall, John Smith was driving the truck at the time of
> the accident, and the plaintiff was a passenger in that truck.
> Ordinarily, in deciding whether the plaintiff was contributorily
> negligent, you would only look at the plaintiff's conduct.

However, there is one situation where John Smith's conduct
affects the plaintiff's ability to recover money. That situation
is where, at the time of the accident, John Smith was the
plaintiff's agent, and was performing duties he was hired to do.

If you find that at the time of the accident, John Smith was the
plaintiff's agent, and was performing duties that he was hired to
do, then any negligence on John Smith's part would transfer to
the plaintiff. It would be as though the plaintiff himself were
negligent.

On the other hand, if you find that John Smith was not the
plaintiff's agent, or that he was not performing duties that he
was hired by the plaintiff to do, then any negligence on John
Smith's part would not transfer to the plaintiff (p. 1351).

It is conceivable that Nagel's (1979) instructions varied in the
degree to which subjects were able to comprehend them, with the 10:1
tradeoff definition being the simplest to comprehend. Alternatively,
comprehensibility might have interacted with the specific techniques
Nagel employed. For example, given the 10:1 tradeoff definition, many
subjects might have considered CI the most undesirable outcome, and
then, as in Nagel's example, placed AG at -10, thereby maintaining a
10:1 ratio. In fact, quantitative definitions (".90 probability," 10:1
tradeoff, etc.) may lead to much greater discrepancies between the
different quantification approaches reviewed above than the
traditional, qualitative definitions do. For example, suppose Nagel
(1979) had employed the self-report approach. Subjects who might find
it very difficult to translate the ".91 probability" definition into

expected utilities, thereby creating a great deal of variance in SDT
estimates, might find it relatively easy to simply mark ".90" on a
zero-to-one self-report probability scale, leading to fewer individual
differences based on perceived costs like the sex differences Nagel
found for the rape case. A close match between quantified
definitions and estimates may result from an artifact of the
measurement process that has little to do with how real jurors form
verdicts.

On the other hand, recent evidence (Anonymous, 1984) suggests
that in some cases, quantified instructions may be more likely than
qualitative instructions to have their intended effect on verdicts,
suggesting legitimate effects upon decision criteria. This issue is
worthy of more systematic attention, since its policy implications are
enormous. For example, in McCullough v. State, the Nevada Supreme
Court recently ruled that the use of quantifed criterion instructions
by a district court judge constituted prejudicial error (Trial,
September, 1983, p. 10). This decision may be counterproductive if,
indeed, quantified instructions function better.

Motivation to comply with judicial instructions. There is some
evidence that (a) jurors may not always obey judicial instructions,
and that (b) some judicial instructions will induce more compliance
than others. Research by Doob and Kirshenbaum (1972), Hans and Doob
(1976), Sue, Smith and Caldwell (1973), and Sue, Smith and Gilbert
(1974) has demonstrated that evidence ruled as inadmissable by the
judge (e.g., prior criminal record) can influence mock jurors'
verdicts, despite the admonishments of the judge to the contrary
(although see Cornish & Sealy, 1973, for evidence of compliance). Wolf

and Montgomery (1977) have demonstrated that a strong admonishment by the judge ("...you have no choice but to disregard [the inadmissable evidence]") can actually induce reactance (cf. Brehm, 1966) in subjects, leading to increased disobedience. Further evidence for this possibility is reported in Broeder, 1959.

It is possible that instructions such as Nagel's "10:1 tradeoff" definition motivate greater compliance in subjects than other instructions. The ".90 probability" definition may make the fact that a guilty defendant can be acquitted based on insufficient evidence especially salient for subjects, leading to resentment and reduced compliance, especially by females in a rape trial. On the other hand, the "10:1 tradeoff" definition may remind subjects why our courts are willing to risk such an acquittal: we wish to avoid the even greater tragedy of convicting an innocent person.

"Rationalization." The model of juror decision-making presented in Figure 1 hypothesizes that jurors (a) form an estimate of the probability that the defendant is guilty, and (b) form a decision criterion, ideally based upon the judge's "reasonable doubt" instructions, but to the extent that these are incomprehensible or jurors are not motivated to comply with them, also based upon the jurors' own perceived costs of juridic errors. At the culmination of a trial, (a) and (b) are combined to form a verdict.

Nagel (1979) reports some evidence suggesting that this normative model may not always correctly describe the decision process. Nagel classified subjects as either conviction- or defendant-prone based upon their subjective expected utility estimates. Some subjects received a ".75 probability" definition of reasonable doubt, and

31

others received a ".90 probability" definition. Nagel reports that while conviction-prone subjects tended to estimate p(G) as greater than .90 in the .90 condition, some estimated p(G) as greater than .75 but less than .90 in the .75 condition. Conversely, while defendant-prone subjects tended to estimate p(G) as less than .75 in the .75 condition, some estimated p(G) as less than .90 but greater than .75 in the .90 condition. Thus, Nagel suggests an alternative process, in which jurors (a) form a tentative verdict, (b) receive a reasonable doubt criterion from the bench, and then (c) adjust their estimate of p(G) so that it is consistent with (a) and (b).

Nagel informed other subjects that a hypothetical defendant had either a .60 or a .80 probability of guilt. He then solicited self-report estimates of the law's reasonable doubt standard. Conviction-prone subjects in the .80 condition tended to provide estimates of the criterion that were greater than .60 but less than .80, while some conviction-prone subjects in the .60 condition provided estimates less than .60. There was also some evidence of a converse pattern for defendant-prone subjects. Thus, Nagel also suggests a third possible process, in which jurors may (a) form a tentative verdict, (b) estimate p(G), and then (c) report a criterion that is consistent with (a) and (b).

Nagel (1979) cautions that "the findings concerning the rationalization phenomenon were not as clearcut as [they are described above]. That description represents a simplification designed to clarify the general tendencies (p. 194)." Note that while Nagel treats these findings as evidence of actual discrepancies between the criterion-setting model and the manner in which actual jurors form verdicts, it is possible that his alternative processes are

32

methodological artifacts. Nagel uses the decision theory estimate of the criterion, formula [13], to classify his mock jurors as conviction- or acquittal-prone, and a second estimate, the self-report approach, to represent their functional decision criteria. Yet, both are presumably estimates of the same construct, and Nagel doesn't explain why they should be used differently. As Dane (1979) reports, the decision theory method provides more accurate estimates than the self-report method. One plausible explanation for Nagel's apparent findings is that the use of the zero-to-one scale on both the instructions Nagel provided and the scales he employed created an artificial decision-process that would not have taken place otherwise.

Modeling the impact of the judge's instructions. None of the aforementioned attempts to model the decision criterion has explicitly dealt with the role of the judge's criterion definition. Conceivably, the judge's definition could influence each juror's criterion level in one of three ways:

A. The judge's instructions might influence the juror's utility estimates, thereby influencing the juror's criterion level;

B. The juror might set a personal criterion level, based on utility estimates (i.e., the SDT model), and then adjust the criterion if it is clearly discrepant from the judge's definition as the juror understands it; for example:

$$p* = (p*'' + \Delta p*), \quad \text{where} \quad \Delta p* = c(p*' - p*'') \quad [14]$$

in which p*' represents the juror's perception of
the judge's ideal criterion, p*" represents the
juror's personal criterion, and ^p* represents the
perceived discrepancy between p*' and p*" multiplied by a
scaling constant between zero and one; or

C. The juror's functional decision criterion may be a
weighted average of the juror's perception of the
judge's criterion definition and the juror's personal
utility-based criterion level; for example:

$$p* = p*' (\phi) + p*" (1 - \phi)$$ [15]

in which p*' represents the judge's ideal criterion,
p*" represents the juror's criterion based upon SDT
formula [12], and ϕ represents a weighting parameter on a
zero- to-unity metric. This parameter might be a multiplicative
function of the juror's ability to comprehend the judge's
instructions, and his/her motivation to comply.

Model A would appear to be the model implicitly subcribed to by
the judicial system. In this model, the judge's charge to the jury
ideally educates them and eliminates their personal biases. This
model predicts that judicial instructions will affect both jurors'
utility estimates and their criterion estimates. Ideally, to the
extent that the judge is able to convey the court's standards for the
relative costs of Type I and II errors, extralegal victim and
defendant characteristics should have no effect on the criterion or
the verdicts.

Model B predicts that judicial instructions have a homeostatic effect -- they define a reference level that each juror will presumably attempt to match. Thus, if a juror's utility estimates yield an extremely lax criterion and then receives a more stringent definition from the bench, then the juror may raise the criterion enough to bring it in line with the judge's level. However, the scaling constant, c, indicates that this adjustment might not be complete. This model predicts that the judicial definition will have no effect on jurors' utility estimates and some effect on jurors' criterion estimates, while victim and defendant characteristics will have an effect upon utility estimates and their effect upon criterion estimates will be independent of the judicial definition.

Model C suggests that the clarity of the judge's definition of reasonable doubt (for a given juror), and the juror's motivation to comply with the judge's definition, both serve as important moderators of the effects that judicial instructions will have on each juror's functional criterion level. Thus, if both clarity and compliance are high, the juror will set approximately the same decision criterion that the judge would set. However, it appears unlikely that this close match will happen consistently. First of all, to the extent that judicial definitions of reasonable doubt are vague (as described above), jurors will have a great deal of discretion to set a criterion level as they see fit. Not only should clarity effect the degree to which jurors rely on the judicial definition, it should also effect their estimate of $p\textbf{*}'$ when they do attempt to rely on that definition. Second, many jurors may choose not to comply with the court's admonishments if they perceive that doing so will prevent them from maximizing their own personal utilities. Thus, for jurors with low

35

levels of p* and p(G), Model C would predict that the judicial
definition will have little effect on either utility or criterion
estimates, while victim and defendant characteristics will effect
both.

From Juror Verdicts to Jury Verdicts

By and large, the vast majority of jury simulation studies have
examined only individual verdicts. Bray and Kerr (1982) surveyed 72
such studies and found that only 52% obtained data from groups and
only 29% used the group as the unit of analysis. The relative lack of
data from juries is understandable given the exorbitant costs of
obtaining sufficiently powerful samples of groups. However, there are
a number of reasons why individual verdicts are extremely informative
by themselves. First of all, they are the single best predictor of
group verdicts (cf. Stasser, Kerr, & Bray, 1982). According to
Grofman (1977, p. 192), "it appears certain that the size of the
predeliberation majority largely determines the verdict outcome." Or,
as Kalven and Zeisel put it in their landmark book, The American Jury
(1966, p. 489): "The deliberation process might well be likened to
what the developer does for an exposed film: it brings out the
picture, but the outcome is pre-determined."

Nevertheless, there are a number of reasons why deliberating
groups were of special interest in the present study. First of all,
the reasonable doubt criterion may have important implications for the
establishment of consensus (e.g., Kerr et al, 1976; Stasser, Kerr, &
Bray, 1982, p. 251). Second, group verdicts will permit an
examination of two hypotheses described in the sections that follow.

Kaplan's Evidentiary Polarization Hypothesis. The use

of deliberating groups in the present study also allows for a

conceptual replication of research by Kaplan and Miller (1978)

which suggests that the process of deliberation increases the weight

(in information-integration theory terminology) jurors place upon

evidentiary information, thereby attentuating the effects of non-

evidentiary information (e.g., attorney obnoxiousness).

Note that this research may appear to be in direct

contradiction to the sizable literature on the group polarization

effect (e.g., Myers & Lamm, 1979; Stasser, Kerr, & Davis, 1981), which

suggests that group deliberation typically polarizes individual

predispositions, as reflected in the predeliberation distribution of

opinions. For example, Bray and Noble (1978) composed six-person mock

juries of either high or low authoritarian subjects and had them

deliberate a murder trial. Prior to deliberation, low authoritarians

recommended significantly lower sentences (M = 38.07 years) than did

high authoritarians (M = 56.36). Deliberation had the effect of

polarizing this difference (M = 28.58, 67.70, respectively). Myers

and Kaplan (1976) had mock juries reach judgments for four high-guilt

and four low-guilt traffic felony cases; each jury discussed two of

each and decided the other four privately. Myers and Kaplan report

polarization effects for both judgments of guilt and for recommended

sentences, but only for cases that were discussed in group

deliberation.

However, Kaplan (Kaplan, 1977; Kaplan & C. Miller, 1977; Kaplan &

L. Miller, 1978) and others (e.g., Anderson, 1981, pp. 386-388) have

interpreted the polarization phenomenon in terms of information-

integration theory (described above). Each juror's extralegal bias

37

is conceived of as a piece of information, with a scale value and weight, which is integrated with evidentiary information in formaing a judgment. Kaplan argues, however, that the latter information will predominate during deliberation. Each juror's post-deliberation judgment, then, will be a weighted average of the non-evidentiary bias with all the information valued and weighed during the deliberation process. Consider a juror with a relatively neutral pre-existing bias, with a scale value of 1, and with an evidentiary fact having a scale value of 6. This juror will have a pre-deliberation judgment falling between 1 and 6, depending on the relative weights applied to the two components. Now assume that this juror's judgment is representative of the jury as a whole, although the evidentiary information that other jurors bring to discussion may not be redundant. If she is exposed to new arguments having the same scale value of 6, her post-deliberation weighted average will approach 6. Thus, adding information of the same scale value can have the seemingly paradoxical effect of polarizing judgment, a phenomenon which Anderson (1981) refers to as a Set-Size Effect. The juror's judgment in such a situation will only remain unchanged if she did not weigh her pre-deliberation bias at all.

Kaplan (1977) provides evidence for this line of reasoning using a bogus note-passing procedure that allowed him to control the content of deliberation. Trial transcripts were constructed to have either an exonerating or an incriminating appearance, and bogus notes were constructed to have either the same or the opposite proportion of pro-conviction to pro-acquittal arguments as the notes each actual juror provided. As predicted, when subjects received notes with the same

38

value that they themselves provided, their judgments polarized in the direction of their initial predisposition.

Kaplan and Miller (1978, Exp. 3) report a study in which extralegal biases were induced by manipulating the degree of obnoxiousness of various trial participants (the prosecutor, defense attorney, judge, or experimenter) as well as the appearance of guilt in order to test the hypothesis that only evidentiary information polarizes. Pre-deliberation judgments supported both the extralegal bias and trial appearance manipulations. However, post-deliberation judgments revealed significant polarization shifts for the appearance of guilt but no significant differences due to extralegal biases.

From a legal and social standpoint, this pattern is encouraging. The polarizing effects of the evidentiary factor are robust and dramatic. However, the magnitude of these effects suggests a possible artifactual interpretation of the lack of bias in post-deliberation judgments. On a 0-21 point scale, the post-deliberation judgments cluster around 15 for high- and 6.5 for low-appearance of guilt. While these ratings are not at the actual ceiling and floor of the scale, they may be near the psychological ceiling and floor, given subjects' general reluctance to use scale extremes. If this is the case, then the lack of biasing effects could be the result of a restriction in range. This possibility is made more plausible by the fact that subjects have been explicitly discouraged from using the extremes of the guilt scale in previous research using this general paradigm (Kaplan & Kemmerick, 1974, p. 496). The present study used a case constructed to fall as close to the midpoint as possible. Thus, it was possible to examine whether any biasing effects due to victim and/or defendant attractiveness were polarized or attenuated by group

39

deliberation. If the criterion-setting model is accurate, extralegal

bias should exert its influence independently of the weight of

evidence, and the set-size effect would not apply. In fact,

any extralegal bias should be free to polarize independently of

evidentiary influences.

The Asymmetry Effect. The present study provided an

opportunity to examine the asymmetry effect often found in jury

research (cf. Stasser, Kerr, & Bray, 1982). Researchers have detected

a consistent "leniency shift," in which the rate of conviction tends

to be lower among juries than among jurors. When Social Decision

Scheme matrices, which illustrate the probability of a jury of every

given pre-deliberation split reaching a given verdict, are plotted,

many studies have found an asymmetry effect, in which jurors who are

intially at a deadlock are more likely to move toward acquittal than

guilt. Factions favoring acquittal are also more successful than

factions of the same size favoring conviction at winning converts and

ultimately prevailing. Of course, group polarization studies like

those discussed above provide occasional exceptions to this pattern;

nevertheless, it appears frequently in mock jury research.

One possible explanation for this effect is that the judicial

norms of "presumption of innocence," "burden of proof," and

"reasonable doubt" make it easier to argue for acquittal than for

conviction during deliberation (cf. Nemeth, 1977). If this is the

case, this shift should be eliminated when jurors receive a "mere

preponderance of evidence" instruction from the bench. The present

study also tested this prediction.

CHAPTER 2

METHOD

Subjects and Design

Four hundred and fifty-two volunteers, 139 males and 313 females,
were recruited from Michigan State University Introductory Psychology
courses. In compliance with Departmental and University standards and
procedures, subjects provided informed consent and received extra
course credit for their participation. Although every effort was made
to recruit equal numbers of males and females in each condition, past
experience has shown that males are considerably more difficult to
recruit, and the present study was no exception.

Early in the duration of the experiment, an unfortunate but
serious typographical error was discovered on the first page of the
individual pre-deliberation questionnaire. Subjects were accidentally
informed that the _minimum_, rather than maximum, sentence for auto
theft, the crime in question, was 20 years imprisonment. This
statement is clearly erroneous, if not outlandish, and probably
elicited a variety of reactions from subjects. Because its extreme
implications are irrelevant to the purposes of the present study, it
was not incorporated into the design as an additional factor.
Instead, the typographical error was corrected, and those subjects who
encountered it were omitted from the analyses presented here. Data
from the 321 subjects, 93 males and 228 females, who received the

41

corrected questionnaire are presented.

A 2 (Victim Attractiveness) X 2 (Defendant Attractiveness) X 2 (Judicial Reasonable Doubt Definition: Mere Preponderance of Evidence vs. Reasonable Doubt) factorial design was employed. In order to ensure that the attractiveness factors were not in any way confounded with the specific stimuli employed, two additional control factors, Specific Victim Photo and Specific Defendant Photo, were also included in the design. These factors are nested within the Victim Attractiveness and Defendant Attractiveness factors, respectively. Cell sizes for the main design are displayed in Table 1:

Table 1: Cell Sizes for the Experimental Design

Instructions

Victim	Reasonable Doubt		Preponderance of Evidence		
	Male	Female	Male	Female	Defendant
Attractive	10	25	13	32	Attractive
	8	27	13	29	Unattractive
Unattractive	12	30	11	31	Attractive
	13	23	13	31	Unattractive

The hypotheses of the present study also required data at the group level of analysis. For this reason, an attempt was made to schedule subjects in groups of four. Ultimately, 236 subjects participated in 59 4-person juries, 33 subjects participated in 11 3-person juries, 34 subjects participated in 17 dyads, and 18 subjects were not able to participate in groups and were not included in analyses at the group level. Thus, at the group level of analysis, the

42

design included a Size factor. Subjects were nested within groups, and groups were nested within the experimental conditions.

Stimulus Materials and Pilot Studies

Attractiveness manipulations. Eight black and white photographs were required -- two attractive males, two unattractive males, two attractive females, and two unattractive females. A pool of 16 male and 10 female black-and-white photographs with good "face validity" were selected from the collections of several departmental researchers. These photographs were originally obtained from a number of sources, including high school and college yearbooks; none of them had been used in research during the year prior to the present study. In an initial pilot study conducted during the term prior to the main study, 12 males and 14 females from the University Psychology Department were recruited to select suitably attractive and unattractive photographs. Participants read the following instructions:

> THANK YOU FOR YOUR PARTICIPATION AND INTEREST IN THIS RESEARCH. We are planning a large, comprehensive program of research on criminals and criminality. We are especially interested in what types of people commit felonies, and in discovering what types of factors influence (a) the probability that they will be convicted of a crime that they have committed, and (b) the probability that they will successfully adjust to the community after prison.
>
> Today's study is a preliminary look at the question: Can people recognize "criminality" in facial photos? We would like you to

43

take a few moments to examine the booklet of 26 facial photos,
and to evaluate these faces. All 26 photos were taken from high
school and college yearbooks. Some of these photos may depict
people who were later convicted of felonies and served time in
federal prisons. Others are ordinary people who have not
committed felonies. Of course, you will not know which are which.
Please fill out the questionnaire for each photo. Do not write on
the photo sheets. Since this is a pilot study (i.e., a
preliminary one), we would find it very helpful if you added
comments and additional impressions in the margins of the
questionnaire. Let us know what you think of each photo.

The photo booklets consisted of 26 photos arranged on three
consecutive 9-1/2" x 14" sheets of paper, with an arbitrary three-
digit I.D. number under each photo. The questionnaire consisted of
26 sets of the following scales:

Photo # _____

```
     Extremely                              Extremely
    attractive :___:___:___:___:___:___:___: unattractive


     Extremely                              Extremely
 unintelligent :___:___:___:___:___:___:___: intelligent


     Extremely                              Extremely
   trustworthy :___:___:___:___:___:___:___: untrustworthy
```

How likely is it, in your opinion, that this person has been, or will
be, convicted of a felony?

```
     Extremely                              Extremely
      unlikely :___:___:___:___:___:___:___: likely
```

This procedure was used to select eight suitable photographs that were

perceived as attractive or unattractive but which were relatively

neutral on the remaining three dimensions. Scale ratings for the eight
photographs selected are presented in Table 2.[3]

Table 2

Pilot Study Scale Ratings

for Victim and Defendant Photographs

Sex	Attractiveness		Intelligence		Trustworthiness		Possible Felon?	
M	1.69	(.79)	3.69	(1.49)	3.46	(1.56)	3.73	(1.69)
M	2.19	(1.39)	3.65	(1.67)	3.62	(1.55)	4.35	(1.50)
F	1.46	(1.24)	5.12	(1.63)	5.15	(1.63)	1.96 *	(1.43)
F	2.28	(1.34)	3.72	(1.75)	3.80	(1.61)	3.44 *	(1.73)
M	5.20	(1.08)	4.84	(1.43)	4.92	(1.61)	3.08	(1.91)
M	5.73	(.96)	4.96	(1.34)	4.92	(1.52)	2.92	(1.70)
F	5.92	(.69)	4.62	(1.24)	5.11	(1.21)	2.15 *	(.83)
F	6.04	(1.10)	4.84	(1.65)	5.72	(1.06)	1.92 *	(1.88)

NOTE: Attractiveness and Trustworthiness scales have been recoded so
that 1 is the low anchor and 7 is the high anchor for each rating
scale. Standard deviations appear in parentheses. Means denoted by
asterisk are not relevant for the present study, as the female
portrayed a felony victim, not a criminal defendant.

Trial Materials. The trial simulation was a modification of a

trial transcript used previously in our lab (e.g., Kerr et al, 1982).

Although the transcript is very realistic (including, for example,

opening and closing arguments, direct- and cross-examination of

witnesses, and judge's instructions to jurors), the case is, in fact,

a fictional one. This permitted evidence to be manipulated in series

of minor pilot studies (approximately ten subjects each) used to establish a close case which would avoid both floor and ceiling effects on verdicts. This attempt was very successful, resulting in a trial scenario with a 52.6% pre-deliberation conviction rate in the main study.

In general, the case involves an auto theft charge that was allegedly tried in Chicago, Illinois. Briefly, the facts of the case are as follows: The victim's car was stolen while she was shopping. The car was recovered during a police raid on a garage in which a number of stolen cars were being repainted. The defendant's fingerprints were found in the car, a number of checks made out to the defendant were found in the garage, and subsequent investigation revealed that the defendant was in a cafe near the place where the the car was stolen at about the time of the theft. The defendant claimed that he had been an employee of the garage, that he had left his fingerprints when he repainted the car, that the checks had been paychecks, and that he had not left the cafe until well after the theft had occured.

In addition to the defendant and victim photographs, the trial transcript also included photographs of the witnesses, attorneys, and the presiding judge. These photographs were retained from the Kerr et al. (1984) study. The defendant and victim photographs varied by condition, and the additional photographs were constant across conditions.

An audio simulation of the trial was also constructed. Graduate students in the Department of Psychology performed the roles of the judge, victim, defendant, attorneys and witnesses. The audio tape was prepared primarily to keep the rate of presentation of trial materials

46

constant for all subjects. Participants were informed that the tape
was not an actual recording of the trial, but that it was hoped that
they would find that it made the trial simulation more involving and
life-like.

In order to manipulate the judge's criterion instructions,
subjects read and heard either "mere preponderance of evidence"
instructions, as typically used for civil trials, or "reasonable
doubt" instructions. These instructions were obtained from sourcebooks
of patterned jury instructions (Reid, 1960a, 1960b). Both sets of
instructions came from cases tried in the State of Michigan and were
selected because they appeared to be approximately equivalent in
length and comprehensibility and were judged to be fairly
representative. All subjects received the following instructions (with
an adaptation for the "preponderance of evidence" condition in
parentheses):

> Now in this phase of the proceedings the Court explains to you
> what the law is that applies to this case. The Information
> charges one offense, the charge is Auto Theft. The statute upon
> which this charge is based is Article III.45.2 of the Revised
> Illinois Penal Code and it reads as follows:

> "Any person who shall knowingly take possession and operate a
> motor vehicle without the knowledge and permission of the person
> holding title to that vehicle shall be guilty of a felony."
> Auto theft, as charged in the Information has been defined as the
> possession of a motor vehicle without the permission of the
> lawful owner of that vehicle. In this case, it is clear that the
> owner of the vehicle in question did not grant any permission to

47

the person or persons who removed it from the location indicated
in the Information...

In dealing with criminal matters there are several particular
rules which apply and which do not apply to civil matters. One
of these is the doctrine of presumption of innocence. A
defendant is presumed to be innocent until his guilt is
established beyond a reasonable doubt (with a preponderance of
the evidence). In accordance with that rule of law, no inference
of guilt may be drawn from the fact that a person has been
arrested and has been placed on trial.

Juries in the "reasonable doubt" condition then received the following
instructions:

No man can be convicted of a crime in this jurisdiction until his
guilt is established beyond a reasonable doubt. A reasonable
doubt is what the words imply, a doubt founded in reason, a
doubt for which you can give a reason, a doubt growing out of the
testimony in the case or the lack of testimony, a doubt which
would cause you to hesitate in the ordinary affairs of life. It
is not a flimsy , fanciful, fictitious doubt which you could
raise about anything and everything. It means a reasonable
doubt. If, when all is said and done, you have such a doubt
about the accused, it is your duty to acquit him. (People v.
Davis, 171 Mich 241, 137 NW 61.)

Instructions for the "preponderance of evidence" condition were
adapted for use in a criminal trial.   These instructions read as follows

(with the original wording in parentheses):

> The burden of proof in this case is upon the prosecution (the
> plaintiffs) to show by a preponderance of the evidence the
> material facts which the State has (they have) alleged in its
> (their) declaration. By a preponderance of the evidence we mean
> simply the greater weight of evidence; in other words, the
> prosecution (the plaintiffs) in this case must produce evidence
> which in your minds carries greater weight than that which has
> been produced against it. (Blaty v. Gray, 217 Mich 531, 187 NW
> 360.)

> Finally, all juries were told:

> Very well. Members of the Jury, the time has come to submit this
> case to you for your deliberations. As I told you, you have
> nothing to do in this case but to determine the guilt or
> innocence of the defendant. I can tell you what the law is but
> you are absolute in the realm of fact.

Procedure

A maximum of 16 subjects participated during any given session.
Subjects scheduled themselves by signing up for a given session, and
the experimenters, four male and four female undergraduates, called
them the evening before a session to confirm the appointment.

The laboratory featured a large rectangular central room with
three smaller rooms on either side. The four smaller rooms in the
corners were used as jury deliberation rooms, whereas the two remaining
middle rooms were left vacant. There were four chairs and a
rectangular table in each deliberation room; the table was flush

49

against the wall with a chair at each end and two chairs on the exposed side. A microphone on a stand was placed on each table.

As subjects arrived at the laboratory, they were asked by the experimenter to have a seat in one of the four deliberation rooms. The experimenter alternated the room assignments between the front two rooms until there were four subjects in each, and then alternated the room assignments for the back two rooms for the remaining subjects. This procedure was followed in attempt to randomize the composition of the juries as much as possible. An attempt was made to create as many 4-person juries as posible given the attendance at a given session. However, if necessary, 2- or 3-person juries were formed, or subjects were seated alone in a room. The sex composition of the groups was allowed to vary randomly. While subjects were seated, they were asked to read the standard departmental consent form and a brief description of the experiment (Appendix A), and to sign the form if they wished to participate. Subjects were allowed to talk to one another while waiting for the session to begin. The door to each deliberation room remained open during the early portion of each session.

When all the subjects were seated, the experimenter distributed the trial transcripts. Deliberation rooms were randomly assigned to a victim attractiveness/defendant attractiveness photo combination prior to each session, and sessions were randomly designated for either the "reasonable doubt" or the "preponderance of evidence" condition, so that all the subjects in a given deliberation room received identical booklets. Tucked in the back of each transcript was a large manilla envelope containing pre- and post-deliberation questionnaires. One of each set of four transcripts was marked with a large red "F" and also contained a group questionnaire; the recipient of this folder was

50

randomly determined and became the jury's foreperson. If there were
only two or three people in a jury, the experimenter made sure that
one of the jurors received the foreperson's folder.

The experimenter informed subjects that all their instructions
would be provided by an audio tape-recording, and then he or she
turned on a tape recorder in the central room; the recording played
through two speakers placed on either end of the central room, so that
it could be clearly heard in each room. The tape recording played
subjects the following instructions:

> Welcome to "The Jury Study." Thank you all for coming today.
> Today, each of you will take on the role of a juror. You will
> read the written transcript of a criminal trial and be asked to
> reach a verdict and make related judgments for the case. This
> case you will consider is called "The People v. William Lambeth."
> William Lambeth was charged with auto theft and tried in Chicago
> in 1974. We chose this actual case so that this study would be
> as realistic as possible. Therefore, we've altered the original
> transcript only slightly. Testimony has been summarized in a few
> places where it could be done without altering its meaning.
> Also, a few portions of the original transcript have been deleted
> altogether; however, these were always clearly unimportant and
> did not bear on the guilt or innocence of the defendant. For
> example, some of the judge's charge to the jury has been deleted.
> In every case the deleted material was redundant with the
> portions which were retained.
>
> We were also able to obtain photographs from the court, from the

police, and from the files of a major Chicago newspaper. These
photographs are included to make the transcript as realistic as
possible, and to give you as much of the information available to
the real jurors as we could. Although we cannot provide you with
a tape recording of the actual trial, you will hear a taped
reenactment of the trial, performed by graduate students at
Michigan State University. We hope this tape will make the
transcript more involving and life-like. As you listen to the
tape recording, please read along in the transcript. Although you
are not actually a juror today, please try to put yourself in the
role of one of the actual jurors.

The trial lasted approximately 38 minutes. At the conclusion of
the trial, subjects were given additional instructions:

Now that you have read and heard the trial, please open the
manilla envelope in the back of your folder and fill out the
questionnaire. Please do not talk to anyone else while you are
filling out the questionnaire. You may find that some of the
questions seem similar to each other or difficult to answer.
Please give careful consideration to each question, and do the
best you can.

When everyone on your jury has completed the questionnaire,
please tuck them back in the folders, and close the door to your
room. Then you may deliberate the case as a jury, and attempt to
reach a unanimous group verdict. One of you will find a large
"F" on your manilla envelope. You will be the foreperson, and we
would like you to fill out the group jury questionnaire for your

jury. The experimenter will notify you when there are only 5
minutes left for deliberation. If you are not able to reach a
unanimous decision at the end of the deliberation period, the
foreperson should indicate that the jury has "hung" on the jury
questionnaire.

At the end of the deliberation period, your experimenter will
open all the doors and sign your experimental credit cards.

When all the members of a given jury completed the pre-
deliberation questionnaire and closed the door to their room, the
experimenter turned on a reel-to-reel tape recorder which recorded
that jury's deliberation. Juries were given approximately 30 minutes
to deliberate, although this period varied widely depending on how
long it took the jury to first complete the pre-deliberation
questionnaires. The average deliberation length was 11 minutes and 47
seconds. At the completion of the session, the experimenter would
sign subjects' experimental credit cards and give them a debriefing
sheet providing some general background on jury research, telling them
how they could contact the principal investigator should they desire
more information about the purpose and/or results of the study, and
requesting their confidentiality until all the experimental sessions
were completed.

Dependent Measures

The Pre- and Post-Deliberation and Group questionnaires appear in
Appendix A. The Pre-Deliberation juror questionnaire consists of 18
items that assessed each juror's pre-deliberation verdict, recommended
sentence, evaluations of the defendant, the victim, and the judge's
instructions, and a series of items intended for use in modeling

jurors' individual decision processes, including subjective expected utilities and subjective probability estimates.

After providing a tentative verdict, jurors rated their confidence in that verdict on a 11-point Likert-type scale. These two measures are required in order to perform the Thomas and Hogue (1976) modeling analyses. In addition, they can be combined into a 22-point Guilt scale, with 1 representing complete confidence in a Not Guilty verdict, and 22 representing complete confidence in a Guilty verdict. This measure has the advantage of being more sensitive than dichotomous verdicts, and can be a valid predictor of movement during group deliberation (cf. Stasser, Kerr, & Davis, 1980).

Subjects were asked to imagine that the defendant was found guilty and they were asked what sentence they would recommend if they were the judge presiding over the case. This type of dependent measure has been criticized (e.g., Konecni & Ebbesen, 1982, pp. 28-29) because real jurors don't make such a judgment in many states. It would clearly be foolish to argue for policy changes on the basis of such data, but the question of how punitive subjects are in response to varying trial conditions is a psychologically valid and meaningful one. In addition, this measure facilitates comparison with past studies.

Subjects rated the victim and the defendant on four 7-point Likert-type items assessing believability, likeability, attractiveness (the manipulation check), and intelligence. Seven-point Likert-type scales were also used to assess how important the judge's instructions were for their decision, how comprehensible the judge's instructions were, and how much they sympathized with the victim and with the

defendant.

Subjects' subjective probability of guilt [p(G)] estimates were assessed using three different measurement formats. First, they rated p(G) by placing a check mark along a 132 millimeter scale ranging from 0, complete certainty of innocence, to 1.0, complete certainty of guilt. Next, they rated p(G) on a 10-point checklist, ranging from "0 chances in 10" to "10 chances in 10" that the defendant did indeed steal the car. Finally, they rated p(G) using an unbounded odds scale. Three different probability formats were used in order to (a) obtain a more reliable estimate of p(G) than a single item can provide, and (b) attempt to establish which format is most accurate and easy for subjects to conceptualize and use.

These three probability formats were also employed to obtain self-report estimates of p*, the minimum probability of guilt the juror requires to render a Guilty verdict. Thus, the accuracy of each format can be assessed by examing whether or not a Guilty verdict is given when p(G) > p*, or a Not Guilty verdict is given when p(G) <= p*.

The Statistical Decision Theory estimate of p* requires subjective estimates of the expected utility of convicting when guilty, convicting when innocent, acquitting when guilty, and acquitting when innocent. Since such estimates are abstract and difficult to quantify, the task was broken down into three steps. First, subjects were asked to imagine each of the four outcomes, one at a time. Next, subjects were asked to consider each outcome and indicate whether they regarded it as a positive outcome or a negative outcome. Finally, subjects were asked to quantify how positive or how negative each outcome would be, using any number ranging from negative

55

to positive infinity.

A final estimate of p% was adapted from Blackstone's comment regarding the relative efficacy of acquitting guilty defendants rather than convicting innocent ones. Subjects were asked to complete the following sentence:

"It is better to let ____ guilty defendant(s) go free than to convict one innocent defendant."

Following deliberation, subjects provided a personal verdict, confidence-in-verdict, their satisfaction with the group verdict, and re-assessed p(G) using the three probability formats. This allowed an assessment of how accurate the p% estimates were at predicting their post-deliberation verdicts. The foreperson was asked to indicate the group verdict on a separate questionnaire.

CHAPTER 3

RESULTS

Manipulation Checks for the Attractiveness Factors

Two items on the pre-deliberation questionnaire assessed subjects' evaluations of the attractiveness of the victim and the defendant. These items were each subjected to a 2 (Subject Sex) x 2 (Instructions) x 2 (Victim Attractiveness) x 2 (Defendant Attractiveness) analysis of variance (ANOVA), presented in Tables B-1 and B-2. As expected, reliable differences between the high and low attractiveness photographs were obtained. The victim was seen as more attractive in the High Attractiveness condition (M = 5.29 on a 7-point scale) than in the Low Attractiveness condition (M = 3.11); $F(1,301)$ = 219.63, $p < .001$. In addition, there was a significant Sex x Instruction x Victim Attractiveness interaction; $F(301)$ = 8.24, $p < .01$. Post-hoc contrasts using the Tukey procedure indicated that the main effect for Victim Attractiveness was reliable ($p < .05$) for both sexes in both instructional conditions. There were no significant sex or instructional differences for the Tukey contrasts at the alpha = .05 level.

Similarly, analysis of the defendant attractiveness ratings revealed a significant main effect for Defendant Attractiveness, $F(301)=303.74$, $p < .001$. As expected, the defendant was perceived as

57

more attractive in the High Attractiveness condition (M = 4.47) than in the Low Attractiveness condition (M = 2.40). In addition, there was a significant main effect for Subject Sex, F(301) = 4.71, p <.05, which was qualified by a Sex x Instructions x Defendant Attractiveness interaction, F(1,301)=13.72, p <.001. Post-hoc contrasts revealed that the interaction was due to a difference between ratings of the attractive defendant by males in the Reasonable Doubt (M = 4.00) and Preponderance of Evidence (M = 5.12) conditions (p<.01). More importantly, the Defendant Attractiveness simple main effects were reliable (p < .01) for both sexes in both instructional conditions.

Recall that not one but two photographs were used to represent the victim and the defendant in each Attractiveness condition. In order to establish that the attractiveness manipulations were not limited to any specific photographs, the victim and defendant attractiveness ratings were each subjected to a 2 (Subject Sex) x 2 (Instructions) x 4 (Victim Photograph) x 4 (Defendant Photograph) ANOVA. A main effect for Victim Attractiveness on the Victim manipulation check was significant, F(1,260) = 89.79, p < .001. T-tests indicated that each pair of High and Low Attractiveness photos was reliably different, t's > 7.64, df > 142, p's < .001. Tukey tests revealed that the High Attractiveness victim photos (M = 5.31, 5.27) did not significantly differ. However, the Low Attractiveness photos (M = 3.65, 2.69) did (p < .01). Additional t-tests were computed to establish that each victim photograph significantly differed from the mid-point (4) of the scale. These one-tailed tests were significant for both attractive photos, t = 9.95, df = 74, p < .001, and t = 10.16, df = 81, p < .001, and both unattractive photos, t = -2.00, df

= 69, p < .05, and t = -9.65, df = 92, p < .001.

A similar pattern was found for the defendant photos.
Decomposition of a main effect for Defendant Attractiveness, F(260) =
100.14, p < .001, indicated reliable differences for each combination
of High and Low Attractiveness photos, t's > 11.33, df > 110,
p < .001. As with the victim photos, Tukey tests indicated that the
High Attractiveness photos (M = 4.52, 4.42) were not statistically
different, but the Low Attractiveness photos (M = 2.09, 2.56) were (p
< .05). Each of the defendant photos was also tested against the mid-
point of the attractiveness scale. These one-tailed tests were
significant for both attractive photos, t = 5.05, df = 58, p < .001,
and t = 4.03, df = 103, p < .001, and both unattractive photos, t =
-12.62, df = 52, p < .001, and t = -12.28, df = 102, p < .001. The
complete pattern of these analyses suggests that the manipulations of
victim and defendant attractiveness were effective.

Establishing the effectiveness of the Instructional manipulation
was more difficult. A number of different estimates of subjects'
criterion for conviction were solicited; however, the validity of
these estimates must be established before they can serve as a
reasonable check on the Instructional manipulation. This is a
particularly thorny issue and will be addressed in more detail later.

Pre-Deliberation Verdicts and Guilt-Related Judgments

Immediately after the completion of the trial simulation and
prior to deliberation, subjects were asked to provide their personal
verdicts. As intended, the case was remarkably close, with an overall
conviction rate of 52.6%. Log-linear modeling techniques (e.g.,
Fienberg, 1970; Knocke & Burke, 1980) were used to assess the impact

of the independent variables upon these dichotomous dependent

measures, and are presented in Table B-3.[4] The Verdict x

Instructions effect was significant, $G^2 = 7.10$, df = 1, p < .01,

shown in Table 3. As predicted, jurors were more likely to convict

the defendant in the Preponderance of Evidence condition than in the

Reasonable Doubt condition. Contrary to predictions, however, there

were no reliable effects for either Victim Attractiveness, $G^2 < 1.00$,

df = 1, or Defendant Attractiveness, $G^2 < 1.00$, df = 1. There were no

significant higher-order effects.

Table 3. Individual Pre-Deliberation Verdicts by Instructions

	Verdict		
	Guilty	Not Guilty	Row Total
Reasonable Doubt	66 (44.6%)	82 (55.4%)	148
Preponderance of Evidence	103 (59.5%)	70 (40.5%)	173
Column Total	169 (52.6%)	152 (47.4%)	321

NOTE: Row percentages in parentheses.

Since verdicts are dichotomous, they may be insensitive to subtle

but real influences upon jurors' judgments. For this reason, a 22-

point "guilt" scale (cf. Kerr, et al, 1982) was created by combining

verdicts with the 11-point confidence-in-verdict scores. Thus, a

guilt score of 1 would represent complete confidence in a Not Guilty

verdict, while a score of 22 would represent complete confidence in a

Guilty verdict. This type of pre-deliberation measure is clearly

related to the verdict and may even add some predictive validity;

Stasser, Kerr, and Davis (1980) report that such confidence scores are related to verdict changes during deliberation. Therefore, the guilt scores were analyzed in a 2 (Subject Sex) x 2 (Instructions) x 2 (Victim Attractiveness) x 2 (Defendant Attractiveness) ANOVA, presented in Table B-4. This analysis replicated the log-linear analysis: The only reliable effect was a main effect for Instructions, $F(1, 305) = 7.27$, $p < .01$. As expected, subjects in the Reasonable Doubt condition were less conviction-prone ($M = 11.09$) than subjects in the Preponderance of Evidence ($M = 13.48$) condition. However, the victim and defendant attractiveness manipulations did not influence pre-deliberation verdicts or verdict-related guilt judgments.

Although the attractiveness manipulations were successful at an aggregate level, not every subject perceived the photographs to be as attractive or unattractive as intended. Furthermore, there were differences in the perceived attractiveness of the photos nested within both the victim and defendant attractiveness conditions. Thus, it was judged reasonable to conduct a number of internal analyses to determine whether there was indeed any extralegal impact of these manipulations for certain photographs or for certain subjects.

First, a 2 (Subject Sex) x 2 (Instructions) x 2 (Victim Attractiveness) x 2 (Defendant Attractiveness) ANOVA was conducted on the guilt scores of subjects who received the most extreme victim and defendant photographs, as described above. This analysis was not successful. There were no reliable attractiveness effects, and the main effect for Instructions was attenuated, $F(1,47) = 3.40$, $p < .075$, presumably due to a loss of statistical power.

Second, a similar 2 x 2 x 2 x 2 ANOVA was conducted on the guilt

scores of subjects who provided an extreme rating (either a 1, 2, 6, or 7) on at least one of the 7-point manipulation checks for attractiveness. This ANOVA is presented in Table B-5. The main effect for Instructions was again attenuated, $F(1,177) = 2.83$, $p < .10$. However, there were two reliable two-way interactions. There was a significant Sex x Victim Attractiveness interaction, $F(1,177) = 4.86$, $p < .03$. However, post-hoc Tukey contrasts revealed no significant simple effects.

An Instruction x Defendant Attractiveness interaction, $F(1,177) = 4.24$, $p < .05$, is displayed in Table 4. Post-hoc Tukey contrasts indicated that the only reliable effect was the simple main effect for Instructions when the defendant is attractive ($p < .01$). It appears that for subjects for whom the attractiveness manipulations were strong, the fact that the defendant was good-looking enabled him to receive the benefit of reasonable doubt where he might otherwise have been convicted. Of course, these internal analyses surrender the advantages of random assignment and are thus only suggestive at best.

Table 4. Instruction x Defendant Attractiveness Interaction
on Guilt Scores for Subjects with Extreme Attractiveness Ratings

	Defendant	
	Attractive	Unattractive
Reasonable Doubt	9.19	12.92
	(37)	(52)
Preponderance of Evidence	14.39	12.81
	(41)	(63)

A 2 (Subject Sex) x 2 (Instruction) x 2 (Victim Attractiveness) x 2 (Defendant Attractiveness) ANOVA was performed on subjects' recommended prison sentences given conviction, and is

presented in Table B-6. There was a marginal main effect for

Instructions, $F(1,286) = 3.85$, $p < .052$. Curiously, subjects were

somewhat more punitive in the reasonable doubt condition (M = 85.25

months) than in the preponderance of evidence condition (M = 72.89).

This may be a logical byproduct of a stricter decision criterion:

jurors who require less evidence to convict may anticipate more post-

decisional regret and subsequently recommend a more lenient

punishment. There were no other significant effects. All in all,

there is little evidence of extralegal bias in these pre-deliberation

judgments.

Evaluations of the Victim and Defendant

In addition to the attractiveness manipulation checks, subjects

also rated the believability, likeability, and intelligence of the

victim and the defendant (where 1 equals the positive anchor and 7

equals the negative anchor) and also indicated their amount of

sympathy for each (where 7 equals maximum sympathy). Each of these

measures was analyzed in a 2 (Subject Sex) x 2 (Instructions) x 2

(Victim Attractiveness) x 2 (Defendant Attractiveness) ANOVA. These

ANOVA's are presented in Tables B-7 to B-14. All interactions were

decomposed using the Tukey procedure.

The Victim. A Sex x Instruction interaction, $F(1,301) = 5.28$, p

< .05, suggests that males found Helen Bednard more credible in the

reasonable doubt condition (M = 1.72) than in the preponderance

condition (M = 2.49), (p < .01). Ms. Bednard was liked more when she

was physically attractive (M = 2.56) than when she was not (M = 3.12),

$F(1, 301) = 16.16$, $p < .001$. Similarly, she was perceived as more

intelligent when she was attractive (M = 2.18) than when she was not

(M = 2.44), F(1,301) = 4.24, p < .04. Decomposition of a Sex x

Instruction x Victim Attractiveness interaction, F(1,301) = 4.34,

p < .04, revealed no reliable comparisons at the Tukey .05 level. An

examination of the grand means for believability, likeability, and

intelligence (M = 2.20, 2.85, and 2.32, respectively) suggests that in

general, the victim was regarded favorably by most subjects. Although

an analysis of the sympathy item indicated a Sex x Instruction x

Victim Attractiveness x Defendant Attractiveness interaction, F(1,299)

= 8.74, p < .003, no post-hoc contrasts were significance.

   The Defendant. Subjects were more likely to believe the

defendant's testimony after receiving the reasonable doubt

instructions (M = 4.14) than after the preponderance of evidence

instructions (M = 4.55), F(302) = 5.44, p < .02, although overall (M =

4.33), subjects were apparently ambivalent about Lambeth's

credibility. Like Helen Bednard, William Lambeth was found more

likeable (M = 3.58), F(1,302) = 4.62, p < .04, and more intelligent (M

= 5.25), F(1,302) = 9.19, p < .003, when he was physically attractive

than when he was not M = 3.87, 5.63, respectively). Thus, ratings of

both the victim and the defendant replicate the well-established

finding that for most people, "what is beautiful is good" (cf.

Berscheid & Walster, 1974). Overall, subjects responded considerably

less favorably to the defendant. Decomposition of a Sex x Victim

Attractiveness interaction, F(1,299) = 5.12, p < .03, revealed no

reliable differences in sympathy for Lambeth, although a Sex x

Instructions x Victim Attractiveness x Defendant Attractiveness

interaction, F(1,299) = 10.40, p < .001, indicated that males in the

reasonable doubt condition sympathized with him more when the victim

was attractive (M = 4.20) than when she was unattractive (M = 2.50), (p < .05). This interaction is unanticipated, and since it does not have any obvious implications for the hypotheses of the present study, will remain uninterpreted.

Pearson product-moment correlations between these evaluative ratings and the predeliberation guilt scores are presented in Table 5. Note that subjects' guilt judgments were more strongly related to their evaluative reactions to the defendant than to their reactions to the victim. Ratings of Lambeth's credibility alone account for about 45% of the variance in guilt ratings. Since Ms. Bednard was not able to positively identify Lambeth as the culprit, his testimony plays a much greater role in the case than her testimony does.

Table 5. Correlations between Evaluative Ratings and Guilt Scores

| Evaluative | Correlations | |
Rating	Victim	Defendant
Believability	-.16 **	.67 ***
Likeability	-.12 *	.29 ***
Intelligence	-.13 **	.07
Sympathy	.13 *	-.43 ***

 * p < .05
 ** p < .01
 *** p < .001

Subjective Probability of Guilt and Criterion Estimates

Self-reported p(G) and p* estimates. Recall that self-reported estimates of p(G) and p* were solicited using three different probability formats, which shall be referred to as the millimeter

65

(MSR), 0 to 10 (TSR), and odds ratio (OSR) self-report methods. These
estimates were all converted to decimal fractions in order to
facilitate comparison on a zero-to-unity metric. Researchers have
speculated (e.g., Kerr, et al, 1984) that such estimates might be
unreliable and difficult for subjects to provide. For this reason, a
multi-trait/multi-method matrix for the two traits (viz., p(G), p$)
and three methods (viz., MSR, TSR, OSR) was constructed to examine
their convergent and discriminant validity (cf. Campbell & Fiske,
1959). This matrix is presented in Table 6.

Table 6. Multi-Trait/Multi-Method Matrix of

Self-Reported p(G) and p$ estimates

		MSR		TSR		OSR	
		p(G)	p$	p(G)	p$	p(G)	p$
MSR	p(G)	---					
	p$	.15 $	---				
TSR	p(G)	.91 $$	.14 $	---			
	p$	-.13 $	.56 $$	-.10	---		
OSR	p(G)	.50 $$	.11 $	.49 $$	-.02	---	
	p$	.08	.25 $$	.07	.34 $$	.56 $$	---

$ p < .05
$$ p < .001

NOTE: Convergent validities appear in boldface type.

Note that the convergent validities are (a) all statistically
significant, (b) higher than the hetero-trait/hetero-method indices,
and (c) higher than the hetero-trait/mono-method indices, thus meeting
Campbell and Fiske's criteria for establishing convergent and

66

discriminant validity. However, the OSR estimates appear to be contaminated by a great deal of method bias, and as a result, do not converge well with the MSR and TSR estimates. Subjects apparently found it more difficult to express uncertainty using the odds format.

Indirect estimates of p*. A rank-ordered (RO) aggregate estimate of p* was calculated in the following fashion. Subjects' mean p(G) estimates were ranked from smallest to largest. Since 152 subjects voted to acquit the defendant prior to deliberation, the 153rd p(G) estimate from the top of the list was found. This value, .55, is the rank-order estimate of the criterion for these jurors, above which a conviction should be obtained.

The Statistical Decision Theory (SDT) estimate of p* was calculated for each subject using formula 12 described above. However, 25 subjects (7.8%) provided "incorrect" valences for at least one subjective utility estimate (i.e., a positive number for UCI or UAG, or a negative number for UCG or UAI). This is consistent with a similar finding by Dane (1979) for 7% (11/168) of his subjects. These discrepancies may reflect genuine values, misunderstanding of the response scale or instructions, or perhaps a lack of sincerity in filling out the questionnaire. If these valences resulted in division by zero or a p* greater than 1.00, no SDT estimate was calculated. All in all, SDT estimates could not be calculated for 45 subjects who either failed to provide one or more utility estimates, provided non-codeable (e.g., verbal) responses, or provided incorrect valences. While the utilities estimates theoretically ranged from negative to positive infinity, responses with an absolute value greater than 9,999,999 were coded as +/- 9,999,999. Any resulting inaccuracies

were judged to be of negligible importance after rounding off
fractions.

"Blackstone" (BLK) p* estimates were obtained by solving for
each subject's response, r, to the statement "it is better to let _r_
guilty defendants go free than to convict one innocent defendant,"
using the formula:

$$p* = \frac{r}{r + 1}$$

As with the SDT utility estimates, r-values exceeding 9,999,999 were
coded as 9,999,999. While subjects were neither encouraged nor
discouraged from providing negative or non-integer values of r, no
subjects did so. Note that the lowest positive value this formula
will yield is .50, provided that subjects use a positive integer for
r. Thus, the BLK p* estimate should be interpreted cautiously, since
subjects who are relatively unconcerned about UCI may have found it
difficult to respond to this item in its present format.

Table 7. Intercorrelations Among p* Estimates

	MSR	TSR	OSR	SDT	BLK
MSR	1.00				
TSR	.56 **	1.00			
OSR	.25 **	.34 **	1.00		
SDT	-.08	.19 **	.14 *	1.00	
BLK	.09	.14	.04	.16 *	1.00

* p < .05
** p < .001

Intercorrelations between the MSR, TSR, OSR, SDT, and BLK

68

criterion estimates appear in Table 7. Since the RO estimate is an

aggregate one, it cannot be included in the correlation matrix.

Unfortunately, attempts to construct a reliable composite index of p*

were unsuccessful; all coefficient alphas were below .70.

Measuring the accuracy of the criterion estimates. "Hit rates"

for each p* estimate were computed in the following manner: If the

p(G) estimate for a given subject was greater than the respective p*

estimate, a "hit" was tallied if the subject's pre-deliberation

verdict was "Guilty" and a "miss" was tallied if the verdict was "Not

Guilty." If the p(G) estimate was less than or equal to the p*

estimate, a "hit" was tallied if the verdict was "Not Guilty" and a

"miss" was tallied if the verdict was "Guilty." Self-report p*

estimates were matched against their respective p(G) estimates, while

SDT and BLK estimates were matched against the mean self-reported p(G)

(coefficient alpha = .84). The RO hit rate was obtained by matching

the aggregate p* estimate of .55 against each subject's mean p(G)

score. Each hit rate was tested against the null hypothesis of 50%.

The average p* estimate, percentage of hits, and z statistic for each

method are presented in Table 8.

Table 8. Mean p* and Accuracy Rates

Estimate	Mean p*	Hits	n	z	prob.
MSR	.68	67%	320	6.08	<.001
TSR	.69	72%	315	7.81	<.001
OSR	.50	70%	307	7.01	<.001
SDT	.54	84%	276	11.30	<.001
BLK	.49	68%	305	6.29	<.001
RO	.55	87%	311	13.09	<.001

The p* estimates are all rather low, ranging from .49 to .69. However, every estimation method was significantly more accurate than expected by chance. The RO and SDT methods, although less direct than the self-report methods, are nevertheless considerably more accurate. It is conceivable, however, that the high accuracy rate for the SDT method might be an artifactual result of the fact that the cases which lead to computational errors were deleted. One might argue that the cases with incorrect utility valences should be tallied as additional "misses," thereby yielding a hit rate of 77%. Therefore, this corrected hit rate was tested against the others. Z-tests indicate that the corrected SDT hit rate is less accurate than the RO hit rate ($z = 3.21$, $p < .001$), more accurate than the MSR ($z = 2.77$, $p < .003$), OSR ($z = 1.93$, $p < .03$), and BLK ($z = 2.48$, $p < .005$) hit rates, but not different from the TSR hit rate ($z = 0.90$, n.s.). This correction seems unreasonably stringent, however. The SDT method did not incorrectly predict those 25 pre-deliberation verdicts; instead, it made no prediction at all.

Alternatively, perhaps the SDT estimate is more accurate because it was only obtained from the most alert, dilligent, or intelligent subjects. Z-tests comparing the mean hit rates for each method, presented in Table 9, suggest that this is not the case. These z-tests were only computed for cases in which both the SDT estimate and the other estimate in question were available. Nevertheless, the SDT estimate was as accurate as the RO estimate, and is significantly more accurate than any of the others (p's $< .001$). The mean MSR, TSR, OSR, BLK and RO hit rates for these selected cases are 67%, 73%, 72%, 67%, and 87% respectively.

Table 9. Z-tests of the Relative Accuracy of p% Estimates

Z Statistic

	SDT	MSR	TSR	OSR	BLK
MSR	-4.85 * (275)				
TSR	-3.35 * (274)	1.38 (314)			
OSR	-3.65 * (270)	1.05 (306)	-0.81 (302)		
BLK	-4.72 * (270)	0.43 (304)	-0.89 (303)	-0.35 (298)	
RO	0.85 (276)	6.09 * (310)	4.70 * (309)	5.15 * (303)	5.64 * (305)

* p < .001   NOTE: Number of subjects per comparison in parentheses.

This pattern of accuracy is consistent with the pattern reported by Dane (1979). Dane also found that the SDT and RO methods of estimating the decision criterion were more accurate than self-reported estimates, with SDT hit rates of 82-85% and RO hit rates of 86-88%.

Z-tests were computed to test for differences in the relative accuracy of each p% estimate at predicting Guilty and Not Guilty verdicts. These analyses only revealed significant effects for the MSR estimate, z = -3.91, p < .001, and the TSR estimate, z = -4.94, p < .001. These estimates were 57% and 60% accurate, respectively, at predicting Guilty verdicts, and 78% and 85% accurate, respectively, at predicting Not Guilty verdicts. Apparently, subjects rendering Guilty verdicts were prone to overestimating their actual decision criterion using either probability format. These two mean p% estimates are considerably higher than the other, less direct estimates, suggesting

that self-presentational concerns may have inflated their criterion beyond its actual level.

Although Iversen (1971) has criticized the use of a zero-to-ten probability format on conceptual grounds, as discussed above, subjects in the present study were nevertheless more accurate with such a format that with the millimeter zero-to-one format. The gain in clarity and accuracy may justify a degree of conceptual murkiness.

Thomas and Hogue Estimates. The Thomas and Hogue model assumes a positive linear relationship between confidence-in-verdict ratings and $|p(G) - p*|$, the absolute difference between jurors' perceived probability of guilt and their decision criterion. This assumption was tested by correlating confidence ratings with absolute difference scores using the MSR, TSR, OSR, BLK, and SDT p* and p(G) estimates. Self-report p* estimates were subtracted from their respective self-report p(G) estimates, and SDT and BLK estimates were subtracted from the mean p(G) index. The self-report estimates provided correlations of -.06, .12, and .09, respectively; only the TSR index was significant (p < .02). The BLK, SDT, and RO estimates yielded correlations of .18, .36, and .41, respectively; all were highly significant (p < .001).

Thus, Thomas and Hogue's (1976) fundamental assumption receives reasonable support from four of the six estimates. As discussed in Chapter 1, these tests of that assumption can only provide independent support for the model to the extent that the p* and p(G) estimates are themselves valid; in this regard, it is encouraging to note that the greatest support was provided by the SDT and RO procedures, the most accurate of the six.

The Thomas and Hogue c and m parameter estimates were calculated

for all subjects using a FORTRAN program which requires verdicts,

confidence ratings, and appropriate contrasts between conditions as

input. As discussed in Chapter 1, these estimates are aggregate, they

have an arbitrary metric, and they have no satisfactory error term.

As a result, they cannot be subjected to inferential statistics and

are therefore only descriptive. In this case, c and m were estimated

for subjects in the Reasonable Doubt and Preponderance of Evidence

conditions. Since subjects in the Reasonable Doubt condition were

less likely to convict the defendant, we would expect the criterion

estimate, $\hat{c}$, to be greater for those subjects than for subjects in the

more lax Preponderance of Evidence condition. However, these

estimates were 1.16 and 1.17, respectively. The respective probability

of guilt parameters, $\hat{m}$, were 1.13 and 1.27. The differences in these

estimates are slight; nevertheless, the complete pattern of Thomas and

Hogue estimates suggests that the instructional manipulation may have

influenced verdicts and guilt scores by shifting subjects' perceived

probability of guilt rather than their decision criteria. This

possiblility is explored below.

Criterion instruction manipulation checks. One-tailed t-tests

were conducted on the MSR, TSR, OSR, SDT, and BLK p% estimates as

planned comparisons of the effectiveness of the instructional

manipulation. Results of these tests are presented in Table 10. If

the instructional manipulation were successful and the p% estimate

were adequately valid and reliable, we would expect a higher criterion

level for subjects in the Reasonable Doubt condition. Although the

pattern of means is consistent with this expectation for five of the

six estimates, the only significant difference was obtained using the SDT estimate, and even this difference is surprisingly small. Again, since the RO criterion is an aggregate estimate, there is no variance to analyze.

Table 10. Instructional Manipulation Checks for Each p* Estimate

| | Means | | | |
p* estimate	Reasonable Doubt	Preponderance of Evidence	df	t
MSR	.67	.68	318	-0.36
TSR	.69	.68	314	0.53
OSR	.51	.49	306	0.83
BLK	.51	.48	313	0.73
SDT	.56	.52	281	1.68 *
RO	.56	.53	---	----

* p < .05

The mean self-report p(G) index for each instructional condition was examined in a oneway ANOVA. As suggested by the Thomas and Hogue analysis, commission was perceived as somewhat less probable (M = .51) for subjects in the Reasonable Doubt condition than for subjects in the Preponderance of Evidence condition (M = .56), although this difference was not significant, $F(1,309) = 3.29$. Thus, the complete pattern of modeling estimates provides only weak support for the predicted criterion shift.

Since the SDT estimate varied as a function of the judge's defined standard of proof, t-tests were conducted to determine whether this resulted from a shift in the expected utility of a specific trial

74

outcome. None of these tests were significant; t values ranged from
-.97 to .57, df = 290 to 298.

Group Verdicts

Because two- and three-person groups were only formed when there
were not enough subjects in attendance to form a four-person group,
only 17 two- and 11 three-person groups were obtained. Unfortunately,
there are not enough groups at either size to allow the use of a
three-leveled Group Size factor in the analysis of group verdicts and
post-deliberation judgments. Following a suggestion by Brown (1981),
the two- and three-person groups were therefore combined and a two-
level, Small vs. Large, Group Size factor was created.

The overall trend for the group verdicts replicates the leniency
bias typically found in mock jury research (Stasser, Kerr, & Bray,
1982). While 52.6% of the individual pre-deliberation verdicts were
for conviction, only 29.9% of the groups voted for conviction, 47%
voted for acquittal, and 23% were unable to reach a unanimous group
verdict.

Log-linear analyses were conducted to examine the effects of
Size, Instructions, Victim Attractiveness, and Defendant
Attractiveness upon the group verdicts. These analyses are presented
in Table B-15. Curiously, the Verdict x Instruction effect obtained
for individual pre-deliberation verdicts was not replicated at the
group level. However, there were reliable effects for Verdict x Size,
$G^2 = 7.88$, df = 2, p < .01, and Verdict x Defendant Attractiveness,
$G^2 = 7.25$, df = 2, p < .01. As shown in Table 11, the Verdict x
Size effect indicates that larger juries were considerably less likely
to reach a unanimous group verdict. This finding conceptually

replicates a similar pattern reported by Kerr and MacCoun (1984).

Table 11. Group Verdicts by Size

	Guilty	Not Guilty	Hung	Row Total
Small	8	18	2	28
	(28.6%)	(64.3%)	(7.1%)	
Large	18	23	18	59
	(30.5%)	(39.0%)	(30.5%)	
Column Total	26	41	20	87
	(29.9%)	(47.1%)	(23.0%)	

NOTE: Row percentages appear in parentheses.

Table 12. Group Verdicts by Defendant Attractiveness

	Guilty	Not Guilty	Hung	Row Total
High	9	26	8	43
	(21.0%)	(60.5%)	(18.6%)	
Low	17	15	12	44
	(38.6%)	(34.1%)	(27.3%)	
Column Total	26	41	20	87

NOTE: Row percentages appear in parentheses.

The Verdict x Defendant Attractiveness effect is portrayed in Table 12. Juries who viewed an attractive defendant were considerably more likely to acquit him than juries who viewed an unattractive defendant. This finding is especially noteworthy because there was no such extralegal bias at the individual, pre-deliberation level. Thus, this finding is completely at odds with Kaplan and Miller's

(1978) contention that group deliberation serves to minimize such extralegal biases by focusing jurors' attention on evidentiary factors.

Table 13. Social Decision Scheme Matrix
for Each Defendant Attractiveness Condition

Initial Split (G, NG)	Group Verdict			Row Total
	Guilty	Not Guilty	Hung	
Attractive Defendant				
0, 4	0 (0)	1.00 (4)	0 (0)	(4)
1, 3	.125 (1)	.625 (5)	.25 (2)	(8)
2, 2	0 (0)	1.00 (4)	0 (0)	(4)
3, 1	.27 (3)	.18 (2)	.55 (6)	(11)
4, 0	1.00 (2)	0 (0)	0 (0)	(2)
Unattractive Defendant				
0, 4	0 (0)	0 (0)	0 (0)	(0)
1, 3	.25 (1)	.75 (3)	0 (0)	(4)
2, 2	.3 (3)	.30 (3)	.40 (4)	(10)
3, 1	.47 (7)	.13 (2)	.40 (6)	(15)
4, 0	1.00 (1)	0 (0)	0 (0)	(1)

Social Decision Scheme matrices (Davis, 1973) were computed for 4-person juries in the attractive and unattractive defendant conditions in order to determine whether the biasing effect was due to differences in the deliberation process. These matrices are presented above in Table 13. Log-linear analyses revealed a Verdict x Initial Split effect, $G^2 = 25.07$, $p < .05$, but no Verdict x Initial Split x Defendant Attractiveness effect. Similar analyses deleting the hung and intially unanimous juries replicated this same pattern. Nevertheless, the table indicates some interesting trends which are discussed in Chapter 4.

Effects of Deliberation on Individual Verdicts

Following deliberation, subjects again provided private verdicts and confidence-in-verdict ratings. Because these individual subjects were nested within groups following deliberation, their post-deliberation guilt ratings are experimentally dependent within groups, and may be statistically dependent as well. Therefore, it was necessary to create mean pre- and post-deiberation guilt scores (cf. Anderson & Ager, 1978) for each group in order to assess the impact of deliberation upon individual judgments. These scores were then analyzed in a 2 (Time: Pre- vs. Post-Deliberation) x 2 (Group Size: Small vs. Large) x 2 (Instructions) x 2 (Victim Attractiveness) x 2 (Defendant Attractiveness) repeated-measures ANOVA, presented in Table B-16. This analysis yielded two significant effects. A main effect for Time, $F(1,71) = 14.44$, $p < .001$, provides further evidence of a leniency shift as a result of deliberation. Overall, subjects leaned toward conviction prior to deliberation ($M = 12.23$) and leaned toward

acquittal afterwards (M = 10.66). Consistent with the group verdicts, there was a significant Time x Defendant Attractiveness interaction, $F(1,71) = 7.48$, $p < .008$. As can be seen in Table 14, this interaction is consistent with the Group Verdict x Defendant Attractiveness interaction described above. Since there was no effect for guilt ratings prior to deliberation, this pattern deviates somewhat from the group polarization hypothesis described in the Introduction. For this reason, the interaction was decomposed using the post-hoc Tukey procedure. Tukey contrasts indicate a significant leniency shift for the attractive defendant following deliberation ($p < .01$). After deliberation, significantly less guilt was attributed to the attractive defendant than to the unattractive defendant.

Table 14. Time x Defendant Attractiveness Interaction
on Individual Pre- and Post-Deliberation Guilt Scores

| | | Defendant Attractiveness | |
		High	Low
Time	Pre	12.19	12.28
	Post	9.47	11.82

Modeling Analyses. Contrary to Kaplan and Miller's prediction, there was a reliable extralegal bias due to defendant attractiveness following deliberation. Because Kaplan and Miller specifically hypothesize polarized evidentiary effects following deliberation, and because their measurement techniques were called into question in Chapter 1, self-reported $p(G)$ estimates were assessed again at the conclusion of the session. Although the criterion-setting model is a model of individual pre-deliberation judgment, we can nevertheless

extrapolate a prediction that the defendant attractiveness

effect for post-deliberation guilt scores and group verdicts should be

mediated by the decision criterion. Time constraints precluded the

assessment of post-deliberation criterion and utility estimates,

measures that usually require more attention, concentration, and

tolerance by subjects. Nevertheless, RO and Thomas and Hogue

estimates could be computed using subjects' final mean p(G) ratings,

verdicts, and confidence scores. Tests of these estimates must be

interpreted with a great deal of trepidation, however. Since

individuals were nested within groups, their verdicts should ideally

be aggregated by group, as they are in the analyses described in the

preceeding section. Unfortunately, there is no satisfactory way to

compute an aggregate dichotomous verdict representing the verdict

choices of the members of a given group, an index that both the RO and

Thomas and Hogue procedures require. Moreover, neither procedure will

yield estimates which can be tested using inferential statistics.

The Thomas and Hogue parameters were computed for subjects in the

Attractive and Unattractive Defendant conditions. As predicted, the

criterion was more stringent for the attractive defendant ($\hat{c}$ = 1.03)

then the unattractive defendant ($\hat{c}$ = 0.90). This is consistent with

rank-order analyses, which provided estimates of .58 and .55,

respectively. Nevertheless, there also appears to be less perceived

weight of evidence against the attractive defendant ($\hat{m}$ = 0.82) than

the unattractive defendant ($\hat{m}$ = 0.92). A 2 (Time) x 2 (Instructions)

x 2 (Defendant Attractiveness) x 2 (Victim Attractiveness) x 2 (Size)

ANOVA, presented in Table B-17, was therefore conducted to see if the

Time x Defendant Attractiveness effect on guilt scores was mirrored by

a similar effect for probability of guilt. The Time x Defendant

Attractiveness interaction was marginally significant, $F(1,68) = 3.16$,

p < .08. Tukey contrasts revealed no significant differences between

means, although there was a trend suggesting a lower probability of

guilt for the attractive defendant (M = .48) than the unattractive

defendant (M = .52) after deliberation, a pattern which is consistent

with the Thomas and Hogue m estimates.

Examination of the Asymmetry Effect

The group and individual post-deliberation verdicts both

demonstrate the leniency bias described by Stasser, Kerr, and Bray

(1982). Deliberation had the effect of making juries more lenient

than jurors, and jurors more lenient after discussion than before

discussion. Social Decision Scheme matrices (Davis, 1973; Stasser, et

al., 1982) were computed to determine (a) whether there was an

asymmetry effect, such that evenly split juries on the first ballot

would be more likely to acquit than convict the defendant, and (b)

whether this effect was moderated by the instructional manipulation.

SDS matrices require all juries to be of the same size; furthermore,

previous theory and research (e.g, Kerr & MacCoun, 1984) indicate that

group process does not follow a simple proportionality rule across

varying small group sizes. Therefore, subsequent analyses will only

include four-person groups.

The complete SDS matrix for all four-person groups is presented

in Table 15. The relationship between initial distribution and final

outcome is statistically significant, $\chi^2 = 24.13$, df=8, p < .003.

Table 15. Social Decision Scheme Matrix

for All Four-Person Groups

Initial Split (G, NG)	Group Verdict			Row Total
	Guilty	Not Guilty	Hung	
0, 4	0 (0)	1.00 (4)	0 (0)	(4)
1, 3	.17 (2)	.67 (8)	.17 (2)	(12)
2, 2	.21 (3)	.50 (7)	.29 (4)	(14)
3, 1	.39 (10)	.15 (4)	.46 (12)	(26)
4, 0	1.00 (3)	0 (0)	0 (0)	(3)
Column Total	.305 (18)	.39 (23)	.305 (18)	(59)

NOTE: Cell frequencies in parentheses.

Several patterns are readily apparent in the matrix. First, as expected, the initial distribution of verdict preferences in a potent predictor of final outcomes for juries reaching unanimous group verdicts. Second, there was a rather high rate of hung juries overall. This has the unfortunate effect of reducing statistical power for the crucial comparisons involving close faction ratios ultimately reaching unanimous verdicts. Nevertheless, asymmetry is apparent in an examination of juries with an initial 2:2 split. These juries had a 50% chance of acquitting, but only a 21% chance of convicting the defendant. Moreover, a comparison of whether the

majority "wins" or "loses" for 1:3 and 3:1 splits, dropping hung

juries, indicates that a three-person faction was more likely to win

if it favored acquittal (80%) than if it favored conviction (71.4%),

= 6.17, df=1, p < .02. Conversely, a minority of one was more likely

to win if it favored acquittal (28.6%) than if it favored conviction

(20%).

In Chapter 1, it was suggested that this asymmetry results from

the reasonable doubt standard and should thus appear for groups in the

Reasonable Doubt condition but not the Preponderance of Evidence

condition. A SDS matrix broken down by the instructional conditions,

is presented in Table 16.

The hypothesized Instructions x Group Verdict x Initial Split

effect was tested in a log-linear analysis. Juries that hung or were

initially unanimous were excluded from the analysis. The 3-way effect

was not significant, G^2 = 3.59, df = 2. An examination of Table 16

suggests that three-person majorities favoring conviction actually

fared somewhat better in the Reasonable Doubt condition. A Verdict x

Instruction chi-square test for independence suggests that this

pattern does not differ from chance expectation, X^2 = 1.53, df = 1. A

comparison of juries with 2:2 initial splits reaching unanimous

verdicts shows that 83% of the Reasonable Doubt but only 50% of the

Preponderance of Evidence juries ultimately voted for acquittal.

Although certainly suggestive, a Fisher's Exact Z test indicates that

this pattern may have resulted from chance, p = .30.

Table 16. Social Decision Scheme Matrix

for Each Instructional Condition

Initial Split (G, NG)	Group Verdict			Row Total
	Guilty	Not Guilty	Hung	
Reasonable Doubt				
0, 4	0 (0)	1.00 (3)	0 (0)	(3)
1, 3	.14 (1)	.57 (4)	.29 (2)	(7)
2, 2	.11 (1)	.56 (5)	.33 (3)	(9)
3, 1	.43 (3)	0 (0)	.57 (4)	(7)
4, 0	0 (0)	0 (0)	0 (0)	(0)
Preponderance of Evidence				
0, 4	0 (0)	1.00 (1)	0 (0)	(1)
1, 3	.20 (1)	.80 (4)	0 (0)	(5)
2, 2	.40 (2)	.40 (2)	.20 (1)	(5)
3, 1	.37 (7)	.21 (4)	.42 (8)	(19)
4, 0	1.00 (3)	0 (0)	0 (0)	(3)

Thus, although the complete matrix demonstrates a leniency bias and an asymmetry effect, the hypothesized role of the criterion instructions in generating these effects was not borne out. In Chapter 4, these results are discussed and interpreted.

CHAPTER 4

DISCUSSION

This dissertation had six objectives. First, it attempted to
replicate and extend previous findings suggesting that the physical
attractiveness of the victim and the defendant of a crime can exert an
extralegal influence upon mock jurors verdicts and/or guilt-related
judgments. Second, several different procedures for estimating
jurors' perceived probability of guilt and decision criteria were
evaluated and compared. Third, the judge's charge to the jury was
manipulated to assess whether the reasonable doubt and preponderance
of evidence standards, as defined in practice, have their intended
influence upon jurors' decision making. Fourth, it sought to test a
model of extralegal bias in juror decision-making which proposed that
jurors' standard of proof mediates the influence of many extralegal
factors, and that this relationship is in turn mediated by the costs
of Type I and II juridic errors. Fifth, the role of group
deliberation in amplifying or possibly attenuating such extralegal
biases was examined. And finally, the hypothesis that asymmetry in
Social Decision Scheme matrices of jury deliberation results from
adherence to the reasonable doubt standard was tested. Each of these
objectives is discussed below.

Contrary to predictions based upon previous research, neither victim nor defendant attractiveness biased the pre-deliberation verdicts or recommended sentences of individual jurors in the present study. These verdicts were reached without detectable extralegal bias and were strongly related to perceptions of the defendant's credibility. Since Lambeth's testimony played a pivotal role in the trial, subjects apparently based their pre-deliberation verdicts primarily on their perception of the evidence. If so, their conscientiousness is laudable, and hopefully representative of the performance of jurors in actual criminal trials.

Nevertheless, it is curious that this study failed to replicate previous research. Several explanations are worth considering. Two potential explanations can be ruled out with confidence. First, the male and female photographs selected for use in the study were clearly perceived as intended. Each pair of attractive and unattractive photographs for each actor were significantly different, and mean ratings of each photograph were reliably far from the neutral point on the scale. Second, floor or ceiling effects are not a plausible candidate for eliminating attractiveness effects. The case was extremely close, with a 52.6% pre-deliberation conviction rate.

Another explanation involves the method and mode of trial presentation. Critics of research on attractiveness and juror judgment (Horowitz & Willging, 1984; Konecni & Ebbesen, 1982) have argued that attractiveness effects may be exaggerated by simulations using otherwise impoverished stimuli. For example:

In the laboratory experiment, defendant's characteristics are
etched in strong relief, which becomes a stark figure on a rather
plain background of trial evidence. In the actual trial,
defendants' characteristics are embedded in a wide and rich
network of evidentiary materials that vitiate the characteristics
to a minor role in the trial's outcome (Horowitz & Willging,
1984, p. 79).

Each of the relevant studies reviewed in Chapter 1 used a written
transcript to simulate a trial. These transcripts ranged from brief
fact sheets (e.g., Efran, 1974) to lengthy and detailed "verbatim"
transcripts with photographs of all the major participants (Kerr,
1978). However, only the present study added an audio re-enactment
which brought the trial to life in "real time." The resultant
increase in information and realism may have been sufficient to drown
out effects due to victim and defendant attractiveness. This
explanation gains credence from the fact that Kerr (1978) obtained
victim attractiveness effects using an auto theft trial transcript
almost identical to the one used in the present study, but without an
audiotape.

A final alternative is related to the previous one, and suggests
that the lack of attractiveness effects may be an unfortunate
byproduct of the mix of audio and written modes of presentation in the
present study. Recall that subjects were instructed to read along
with the written transcript as they listened to the audio re-enactment
of the trial. Indeed, the audiotape was actually included as an
afterthought -- with the primary intention of pacing jurors so that
they would complete the trial simultaneously. The photographs were

mounted on pages of the written transcript, at the point at which each character was introduced. Subjects who kept pace with the audiotape would therefore only view each photograph for the duration of the trial that was transcribed on that page. As a result, these subjects would not have an opportunity to view the photographs at a leisurely pace, as they might have had in previous studies. However, in a real criminal trial, jurors do view the victim and the defendant for an extended period. Thus, the present procedure may have artificially restricted subjects' attention to attractiveness; i.e., rather than artificially augmenting the impact of attractiveness, this study may have artificially diminished it. However, the presence of clear effects of the photographs upon ratings of attractiveness, likeability, intelligence, and sympathy indicates that subjects were at least to a minor extent aware of and influenced by the photographs.

Each of these explanations are clearly ad hoc and speculative. Nevertheless, this study's failure to replicate previous attractiveness effects suggests that caution is warranted in interpreting previous research on the topic. Future research examining the biasing effects of attractiveness should ideally use a videotaped trial simulation.

Regretably, the absense of extralegal bias in these pre-deliberation verdicts does not permit tests of the relevant components of the criterion-setting model discussed in Chapter 1.

Estimates of Perceived Probability of Guilt and the Decision Criterion

This study extended previous research (e.g., Dane, 1979; Simon, 1967; Simon & Mahan, 1971) attempting to provide quantitative

estimates of p(G) and p*, two parameters theorized to be of paramount importance in the legal decision process. A wide variety of specific measurement techniques were adopted, including two different probability formats, an odds format, Blackstone's tradeoff, Statistical Decision Theory, and Thomas and Hogue modeling. This breadth of procedures is noteworthy on both theoretical and pragmatic grounds. Theoretically, several of the methods (e.g., SDT, BLK, Thomas & Hogue) make precise assumptions about the nature of the decision process. Pragmatically, using a wide variety of methods enhances the likelihood that subjects can find at least one format that allows them to access and assess their own cognitive processes.

Self-Report Estimates. The three self-reported estimates of p(G) had good convergent and discriminant validity, and subsequently, their composite index was internally consistent. This is fortunate. Unlike previous research, which has relied upon a single p(G) estimate, the present study can therefore provide more confident estimates of the relative accuracy of the p* estimates that fall on a zero-to-unity scale.

The self-reported p* estimates also demonstrated some convergent and discriminate validity, although less so than the p(G) estimates. These estimates were each more accurate than expected by chance, with hit rates of 67-72%. The millimeter and zero-to-ten formats provided the highest mean p* estimates. Since each was prone to overestimating the frequency of acquittals, it seems likely that these p* estimates are inflated, perhaps by a social desirability bias or good intentions that weren't followed.

Indirect Estimates. The BLK, SDT, rank-order, and Thomas and Hogue estimates each share the mixed blessing of opacity; i.e., they

89

do not directly solicit standards of proof. This blessing is mixed because, while they are less vulnerable to the inflationary influences of social desirablity or rationalization, they may also be less likely to tap the actual ongoing cognitive process. In this regard, it is encouraging to note that the rank-order and SDT methods were the most accurate of all, with hit rates of 88% and 85%, respectively. Dane (1979) reports almost identical accuracy rates using these procedures; thus, these findings appear to be stable. It is not particularly surprising that the RO estimate is so accurate; the procedures used to compute the RO p* estimate and the RO hit rate are both based on the positive monotonic relationship between verdicts and p(G).

The high accuracy rate for the SDT procedure supports the decision theoretic conceptualization of the judgment process (e.g., Fried, K. Kaplan, & Klein, 1975; Kerr, Bull, MacCoun, & Rathborn, 198x). Mock jurors do appear to weigh the utilities of potential trial outcomes in setting their criterion for proof. However, as a measurement procedure, the SDT model has several shortcomings. First, subjects seem to find it difficult and time-consuming to explicitly quantify the necessary utilities. Second, for whatever reason, subjects may not provide adequate data for computing p*.

On the other hand, the RO procedure carries no clear theoretical baggage, for better or worse. But it does have the advantage of being extremely easy to compute. Given verdicts and p(G) estimates, which every mock jury study can easily and quickly solicit from subjects, a very accurate estimate of p* can be computed. Separate estimates can be computed for subjects in each cell of an experimental design. Unfortunately, as a single aggregate point estimate, it has no

variance and cannot be submitted to correlational or inferential statistical procedures.

The Thomas and Hogue c and m parameters have the same problem. Nevetheless, it also is easy to collect the verdicts and confidence-in-verdict ratings the model requires. The present study provided adequate support for the presumed positive linear relationship between confidence ratings and $|p(G) - p*|$ (in Thomas and Hogue's notation, $|X - c|$). Using the two most accurate $p*$ estimates, this correlation was estimated at .36 to .41. Furthermore, the c and m parameters approximately mirrored the SDT and RO $p*$ and mean $p(G)$ estimates.

Compliance with Standard of Proof Instructions

At the completion of the trial, subjects received instructions requiring them to convict the defendant if and only if they perceived that the weight of evidence presented in the case surpassed a recommended criterion. For some subjects, this was the "beyond a reasonable doubt" standard, the common law convention for a criminal trial. Other subjects received a "preponderance of the evidence" standard, a more lax criterion typically reserved for civil disputes in which the State is merely an arbitrator and has no inherent interest in the outcome of the trial.

As predicted in Chapter 1, and as intended by the legal system, jurors were less likely to convict the defendant when given the more stringent reasonable doubt criterion. This pattern was found for both individual verdicts and for the verdict-based guilt scale. This result is consistent with a similar result reported by Kerr, et al. (1976) using reasonable doubt definitions of varying stringency.

91

This instructional effect suggests that jurors required less evidence to convict the defendant when they received the preponderance of evidence condition. However, this prediction received mixed support. Only five of the seven estimates of the decision criterion showed such a pattern; and of the five estimates for which inferential statistics can be computed, only the SDT criterion showed a significant difference. The Thomas and Hogue m estimate suggested that subjects perceived less weight of evidence in the reasonable doubt condition, although this pattern was not significant for the mean p(G) estimate.

Note that the mean p* estimates were fairly low, overall, in the range of .49 to .69. These estimates are in the range prescribed by the preponderance of evidence standard, despite the fact that reasonable doubt is the default standard for a criminal trial. Given the fact that the case was an extremely close one, with a mean p(G) of .54, slight differences in p* estimates could result in significant differences in a discrete variable like the verdict. And given some internal inconsistency in the mean p(G) estimates, and "miss" rates of 22-33% for the p* estimates, it is plausible that real differences could exist and yet fail to be detected.

It is also surprising that the judge's defined criterion manipulation did not have a subsequent influence upon the verdicts of deliberating juries, especially since Kerr, et al. (1976) found differences in group verdicts using a theoretically more restricted range of criterion definitions. However, in the present study, 54.2% of the Preponderance of Evidence juries and only 29.4% of the Reasonable Doubt juries convicted the defendant, a strong trend in the expected direction. Note that this is a 24.8% difference; a

92

difference of only 14.9% in individual pre-deliberation conviction rates was statistically significant. Apparently, the instructional effect failed to reach significance because of a loss in statistical power at the group level of analysis.

The prevalence of hung juries in the present study may have obscured real effects due to criterion instructions. Kalven and Zeisel (1966) report a 5% hung jury rate in their large survey of actual juries. In the present study, 30.5% of all four-person groups, 34.6% of the reasonable doubt juries and 27.2% of the preponderance of evidence juries, failed to reach a unanimous verdict. In an actual trial, a hung jury presumably protects the defendant and often results in dismissal of the case. In this study, one might therefore argue that the higher rate of hung juries for the reasonable doubt condition, in conjunction with the lower rate of convictions for reasonable doubt juries, 19% vs. 39%, constitutes evidence that the reasonable doubt criterion serves to protect the defendant.

However, there is some indication that the hung jury rate is inflated spuriously. Subjects in the study were told to deliberate until they had either reached a unanimous group verdict or exhausted the time available in the session. This admonishment was repeated on the foreperson's instruction sheet. Nevetheless, several experimenters reported that when they would enter a deliberation room to conclude a session, occassionally a group would report that they had "hung a long time ago," and spent their remaining time waiting behind the closed door and perhaps discussing matters unrelated to the experiment. Since juries met in private, this couldn't be prevented. Perhaps a "dynamite charge," i.e., an admonishment to continue

deliberating, would have resulted in unanimous verdicts for many of these groups. It isn't clear whether such deadlocks were related to the instructional manipulation or modal verdict preference. If these deadlocks are premature and simply resulted from the random distribution of some unmotivated yet influential subjects, it would have the effect of obscuring real trends. Conversely, juries that hung by running out of time while still deliberating may have eventually reached unanimity. Thus, the data on hung juries in this study is difficult to interpret.

Extralegal Defendant Attractiveness Bias Following Group Deliberation

Kaplan and Miller (1978) have argued and presented evidence that the deliberation process may attenuate extralegal biases found in individual pre-deliberation verdicts. They suggest that such attenuation is inherent in the public act of deliberating. Jurors are unlikely to raise attractiveness as an issue during deliberation, and their colleagues are likely to discourage such a topic if it arises. Instead, juries are hypothesized to focus predominantly upon the facts. The net result is that the bias component in each juror's judgment comes to weigh less and less as deliberation proceeds.

The pattern of both group and individual post-deliberation verdicts in the present study is in direct contradiction to this argument, however. In this study, juries were significantly less likely to convict the defendant of auto theft when he was physically attractive. There was a significant leniency shift for individual guilt ratings in the attractive defendant condition as a result of group deliberation, and this resulted in a significant difference

94

between final guilt ratings in the attractive and unattractive

defendant conditions. Thus, deliberation brought out a clear

extralegal bias that was not apparent in the judgments reached

privately by individual jurors.

This is not the first study that has found extralegal group

verdict effects which weren't manifested at the pre-deliberation

individual level. Hans and Doob (1976) conducted a mock jury study to

examine whether jurors complied with a judge's instructions to

disregard the defendant's prior criminal record. Subjects read a

transcript of a burglary case, and half were informed that the

defendant had been previously convicted of burglary. Prior to

deliberation, 45% of the jurors in the prior record condition and 40%

of the jurors in the no record condition voted for conviction, a

slight but non-significant trend for extralegal bias. But after

deliberating the case, 40% of the prior record juries convicted the

defendant, while none of the no record juries did. This is a

statistically significant difference (p < .01). Furthermore, Hans and

Doob (1976) recorded 71 comments regarding prior record during the

deliberation of the prior record groups; only 14 of these comments

suggested that the record should not be held against the defendant.

Contrary to Kaplan and Miller (1978), subjects apparently had few

qualms about blatantly discussing extralegal information, even though

they were instructed to ignore such information in reaching their

verdicts.

The results of the present study can be interpreted as an example

of group polarization. This phenomenon, strictly interpreted, would

suggest that slight initial differences in guilt ratings for the

defendant attractiveness manipulation would be amplified by group

discussion, shifting toward greater leniency in the attractive defendant condition and less leniency in the unattractive defendant condition. However, an examination of Table 13 in the last chapter refutes this pattern. First, there was no significant shift in the ratings for the unattractive defendant, and the trend suggests _more_ leniency, not less. Second, guilt ratings for the attractive defendant are initially on the guilty side of the midpoint of the scale, and subsequently shift across that point in the direction favoring acquittal.

This strict interpretation of group polarization deserves some qualification, however. Although the numerical midpoint on the 22-point guilt scale is 11.5, the functional psychological midpoint is almost certainly higher. The leniency bias for guilt-related judgments suggests that a 50:50 pre-deliberation split will _not_ result in a 50:50 post-deliberation split -- on the average, the conviction rate will decrease significantly, as it did in the present study. Furthermore, there is evidence that the effects of attractiveness on verdicts are probably due to special treatment of the attractive individual, not mistreatment of the unattractive individual (Sigall & Ostrove, 1975). Thus, the mean pre-deliberation guilt rating for the unattractive defendant, 12.28, is probably at or very near the functional midpoint, and would not be expected to move toward greater guilt. The attractive defendant's rating starts below this point, at 12.19, and moves to 9.47, a significant polarization effect. Of course, group polarization is a description of data; it does not constitute an explanation or define a psychological process.

One possible explanation for the attractiveness bias is that

deliberation created a shift in the reasonable doubt criterion. In the present study, post-deliberation p% estimates suggest that jurors in the attractive defendant condition had more stringent standards of proof than did subjects who saw an unattractive defendant. Unfortunately, the RO and Thomas and Hogue estimates do not allow a conclusive test. Nevertheless, this is a viable hypothesis which implies a very different judgmental process from the information integration model, as outlined by Kaplan and his colleagues. While their model suggests that bias is integrated with evidence in reaching a verdict, the model advocated in this paper argues that bias is reflected in the setting of a decision criterion. This decision criterion is then matched against the perceived weight of evidence to reach a verdict. Since these two components are not integrated, the "set-size" phenomenon is not relevant, and either component may polarize.

There is also a trend suggesting that the weight of evidence may have also shifted after group discussion, but in a direction opposite to the criterion shift. This might be interpreted as an indication that attractiveness was averaged with the evidence, as in the information integration weighted average model. This is possible, but clearly at odds with Kaplan and Miller's (1978) contention that the evidence overwhelms the biased predisposition following deliberation. This pattern could also result if jurors were to apply the decision criterion, already influenced by their personal reactions to the defendant, to each item of evidence independently, rather than to the evidence as a whole (Cullison, 1977). Consistent with this reasoning, Hans and Doob (1976) report that their prior record juries felt that the evidence against the defendant was stronger, and discussed the

97

most incriminating facts more, than the no record juries. If jurors

do use the criterion in this manner, the model and operations

advocated here must be revised. For example, the method of creating

expected verdicts to assess accuracy would probably result in inflated

hit rates. It would be very difficult to determine whether jurors did

use the criterion in such a piecemeal fashion, however. One method

might be to get independent ratings of the evidence, either piecemeal

or as a whole, from a control group that receives no attractiveness

information at all, or receives attractiveness information at the

conclusion of the trial. Or, jurors exposed to different photographs

could be asked to rate each piece of evidence as it was received. The

information integration model would predict that any effects of

attractiveness upon p(G) should gradually diminish as the evidence

increases, even prior to deliberation. The "piecemeal" criterion

model would not predict any such attenuation; once the criterion was

set, the favored actor would continue to be perceived through rosy

lenses.

The attractiveness biasing effect might be explained in terms of

group process. Although log-linear analyses did not indicate a

significant relationship between defendant attractiveness, initial

verdict distribution, and group verdicts, an examination of Table 13

in the preceding chapter does reveal several suggestive trends.

First, juries in the attractive defendant condition were more likely

to begin with a unanimous preference for acquittal, which might result

from a weak pre-deliberation atttractiveness effect combined with the

vagaries of random assignment. Second, although there is an asymmetry

between the 3,1 and 1,3 juries for both attractiveness conditions,

factions favoring acquittal in the 2,2 juries were more successful --
winning every time -- when they were arguing for an attractive
defendant. This pattern raises the possibility that the defendant's
attractiveness served as a "tie-breaker." Jurors might have earnestly
attempted to discuss the facts of the case, but given such equivocal
evidence, might have ultimately resorted to extralegal cues like
physical attractiveness. For example, "I don't know -- he just looks
too wholesome to steal cars." Corroborating evidence for this "tie-
breaker" hypothesis must await a systematic content analysis of the
deliberation tapes.

Standards of Proof and the Asymmetry Effect

In this study, as in previous research (cf. Stasser, Kerr, &
Bray, 1982) juries were more lenient than might be predicted by
jurors' pre-deliberation verdicts, and jurors were more lenient in
their final private verdicts. Overall, this resulted in an asymmetry
in the final outcomes of four-person juries that started with even
faction sizes -- only 21% ultimately convicted the defendant, while
50% acquitted him. Furthermore, a minority of one was more likely to
convert his or her colleagues if he or she favored acquittal.

This study provided an opportunity to test one possible
explanation of this asymmetry effect, which suggests that the
reasonable doubt criterion used in criminal trials makes for an uphill
climb for factions favoring conviction. On the other hand, factions
favoring acquittal need merely create a reasonable doubt in their
opponents to convert them. In the present study, the manipulation of
standard of proof instructions should have only given this advantage

99

to pro-acquittal factions in the reasonable doubt condition. Results indicated a trend suggesting that juries with an initial 2:2 split acquit the defendant more frequently when they received reasonable doubt instructions, but this analysis involved only 10 groups and was not statistically significant. The data did not support the hypothesized Instruction x Initial Split x Verdict interaction.

However, several aspects of the present study weakened the strength of any tests of this prediction. As discussed above, there was an inflated, and possibly artifactual, hung jury rate which may have obscured important patterns at the group level. Furthermore, when 59 groups are distributed across three verdict options and five initial splits, a total of 15 cells, there is a great reduction in statistical power for the crucial comparisons between close factions reaching unanimous decisions. Also, the unfortunate necessity of using four- rather than six- or twelve-person juries produces only one absolute majority-to-minority ratio, 3:1. This is unfortunate because social psychologists have long established that a minority faction of one has unique psychological properties (Asch, 1956; see Kerr & MacCoun, 1984 for a recent example). In the present case, it may obscure important differences in minority influence. For example, perhaps a minority faction favoring acquittal only noticeably benefits from the reasonable doubt standard when it has more than one member; the advantage may be outweighed by the extreme disadvantage of a lack of social support.

A more direct test of the standard of proof hypothesis for the asymmetry effect could be provided by an experiment in which juries receive either the reasonable doubt or the preponderance of evidence instructions, and then are "stacked" -- i.e., explicitly constructed

to be evenly split on the first ballot. This would focus a great deal more statistical power upon the crucial comparisons for the hypothesis. Of course, there are still other viable explanations for the asymmetry effect, including the possibility that the effect is limited to or pronounced in juries composed of predominantly middle-class college students in their late teens and early twenties. An experiment which "stacked" mock juries composed of either college students or members of a jury pool into even initial splits would provide a direct test of this latter hypothesis. The issue of mock jury composition and external validity are discussed below.

The Mock Jury Technique: Is It Externally Valid?

The use of mock juries, especially juries composed of college undergraduates, is not without its critics (e.g., Konecni & Ebbesen, 1982). Can the decisions reached by students after exposure to a hypothetical trial tell us anything about actual decisions reached by actual juries? A complete review of the issues involved in this question is beyond the scope of this paper; for a thorough review and a persuasive defense of the mock jury simulation strategy, the reader is referred to Bray and Kerr (1982). Nevertheless, a few points should be addressed here.

First of all, there is no way in which the present study could have been conducted using real, deliberating juries. In fact, it would be illegal -- "jury tampering" is a felony. The preponderance of evidence instructions are unacceptable for use in an actual criminal trial. And correlational research would not provide the necessary control gained through random assignment and the ability to hold trial

materials constant. As Bray and Kerr (1982) point out, field research often sacrifices the potential for sound causal inferences afforded by the experimental simulation strategy.

Second, there is no clear a priori reason why the results in the present study would fail to generalize to real juries in real trials. While we can generate a list of obvious differences between this simulation and a real trial -- no voir dire, a more homogeneous jury pool, an abbreviated trail and deliberation period, no real judge, no real outcome at stake, etc. -- none implies an explicit a priori reason why these simulation results should not apply. We might generate hypotheses, but these will be empirical questions, and require data to provide answers. As an example, Kerr, Bull, MacCoun, and Rathborn (1984) asked whether British mock jurors undergo a different decision process than American mock jurors -- an empirical question. They found that the decision processes were the same in two different nations. Are American college students more similar to British college students than to American blue collar workers or retirees on a real jury, or more different? Again, an empirical question. And as the questions become concrete, theories begin to germinate.

Note that the present study makes no policy recommendations for the legal system. Rather, it is an exercise in theory construction. It is the theories developed over the course of many mock jury experiments that will make predictions about actual trial situations, not the point estimates and test statistics obtained along the way (cf. Mook, 1983). In the meantime, the results reported here should best be intepreted:

102

...as 'demonstrations' that may reveal that assumptions inherent

in the law do not always hold or that the legal system works in a

way other than officially prescribed (Davis, Bray, & Holt, 1977,

p. 327).

In other words, psychologists can and should point out potential cracks

in Justice's blindfold, and flaws in her scales; to look is not to touch.

References

Asch, S. (1956) Studies of independence and conformity: I. A minority of one against a unanimous majority. Psychological Monographs, 70, No. 9.

Anonymous. (1984) Legal vs. quantified definitions of standards of proof. Manuscript under editorial review, Law and Human Behavior.

Anderson, L., & Ager, J. W. (1978) Analysis of variance in small group research. Personality and Social Psychology Bulletin, 4, 341-345.

Anderson, N. H. (1981) Integration theory applied to cognitive responses and attitudes. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), Cognitive responses in persuasion, (pp. 361-398). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Berscheid, E., & Walster, E. (1974) Physical attractiveness and heterosexual attraction. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, 7. New York: Academic Press.

Bray, R. M., & Kerr, N. L. (1982) Methodological considerations in the study of the psychology of the courtroom. In N. L. Kerr & R. M. Bray (Eds.), The psychology of the courtroom (pp. 287-324). New York: Academic Press.

Bray, R. M., & Noble, A. M. (1978) Authoritarianism and decisions of mock juries: Evidence of jury bias and group polarization. Journal of Personality and Social Psychology, 36, 1424-1430.

Brehm, J. (1966) A theory of psychological reactance. New York: Academic Press.

Broeder, D. (1959) The University of Chicago jury project. Nebraska Law Review, 38, 744-760.

Brown, M. B. (1981) Two-way and multiway frequency tables -- Measures of association and the log-linear model. In W. J. Dixon (Ed.), BMDP statistical software, (pp. 143-206).

Campbell, D. T., & Fiske, D. W. (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Champagne, A., & Nagel, S. (1982) The psychology of judging. In N. L. Kerr & R. M. Bray (Eds.), The psychology of the courtroom. New York: Academic Press.

Charrow, R. P., & Charrow, V. R. (1979) Making legal language understandable: A psycholinguistic study of jury instructions. Columbia Law Review, 79, 1306-1374.

Cornish, W. R., & Sealy, A. P. (1973) Juries and the rules of evidence. The Criminal Law Review, 208-223.

Cullison, A. D. (1977) The model of rules and the logic of decision. In S. S. Nagel (Ed.), Modeling the criminal justice system (pp. 225-246). Beverly Hills, Calif.: Sage Publications, Inc.

Dane, F. C. (1979) Quantifying the reasonable doubt criterion. Unpublished doctoral dissertation, University of Kansas.

Dane, F. C., & Wrightsman, L. S. (1982) Effects of defendants' and

victims' characteristics on jurors' verdicts. In N. L. Kerr & R.

M. Bray (Eds.), The psychology of the courtroom (pp.83-115). New

York: Academic Press.

Davis, J. H. (1973) Group decision and social interaction: A theory

of social decision schemes. Psychological Review, 80, 97-125.

Davis, J. H., Bray, R. M., & Holt, R. W. (1977) The empirical study

of decision processes in juries: A critical review. In J. L.

Tapp & F. J. Levine (Eds.), Law, justice, and the individual in

society: Psychological and legal issues. New York: Holt.

Doob, A. N., & Kirshenbaum, H. M. (1972) Some empirical evidence on

the effect of s.12 of the Canada Evidence Act upon an accused.

Criminal Law Quarterly, 15, 88-96.

Efran, M. G. (1974) The effect of physical appearance on the judgment

of guilt, interpersonal attraction, and severity of recommended

punishment in a simulated jury task. Journal of Research in

Personality, 8, 45-54.

Elwork, A., Sales, B. D., & Alfini, J. J. (1977) Juridic decisions: In

ignorance of the law or in light of it? Law and Human Behavior,

1, 163-189.

Feinberg, S. E. (1970) The analysis of multidimensional contingency

tables. Ecology, 51, 419-433.

Feinberg, W. (1971) Teaching the Type I and Type II errors: The

judicial process. The American Statistician, 25, 30-32.

Fried, M., Kaplan, K. J., & Klein, K. W. (1975) Juror selection: An analysis of voir dire. In R. J. Simon (Ed.), The jury system in America: A critical overview. Beverly Hills, Calif.: Sage Publications, Inc.

Grofman, B. (1977) Jury decision-making models. In S. S. Nagel (Ed.), Modeling the criminal justice system (pp. 191-204). Beverly Hills, Calif.: Sage Publications, Inc.

Grofman, B. (1981) Mathematical models of juror and jury decision-making: The state of the art. In B. D. Sales (Ed.), Perspectives in Law and Psychology. Volume 2: The trial process. New York: Plenum Press.

Hans, V. P., & Doob, A. N. (1976) Section 12 of the Canada Evidence Act and the deliberations of simulated juries. Criminal Law Quarterly, 18, 235-254.

Horowitz, I. A., & Willging, T. E. (1984) The psychology of law: Integrations and applications. Boston: Little, Brown and Co.

Iversen, G. R. (1971) Operationalizing the concept of probability in legal-social science research. Law and Society Review, 5, 331-333.

Izzett, R., & Leginski, W. (1974) Group discussion and the influence of defendant characteristics in a simulated jury setting. Journal of Social Psychology, 93, 271-279.

Izzett, R., & Fishman, L. (1976) Defendant attractiveness as a
function of attractiveness and justification for actions.
Journal of Social Psychology, 100, 285-290.

Kalven, H., & Zeisel, H. (1966) The American jury. Boston: Little,
Brown.

Kaplan, J. (1968) Decision theory and the factfinding process.
Stanford Law Review, 20, 1065-1092.

Kaplan, M. F. (1977) Discussion polarization effects in a modified
jury decision paradigm: Informational influences. Sociometry,
40, 262-271.

Kaplan, M. F. (1982) Cognitive processes in the individual juror. In
N. L. Kerr & R. M. Bray (Eds.), The psychology of the courtroom
(pp. 197-220). New York: Academic Press.

Kaplan, M. F., & Kemmerick, G. D. (1974) Juror judgment as
information integration: Combining evidential and nonevidential
information. Journal of Personality and Social Psychology, 30,
493-499.

Kaplan, M. F., & Miller, C. E. (1977) Judgments and group discussion:
Effect of presentation and memory factors on polarization.
Sociometry, 40, 337-343.

Kaplan, M. F., & Miller, L. E. (1978) Reducing the effects of juror
bias. Journal of Personality and Social Psychology, 36, 1443-
1455.

Kerr, N. L. (1978a) Beautiful and blameless: Effects of victim

attractiveness and responsibility on mock juror verdicts.

Personality and Social Psychology Bulletin, 4, 479-482.

Kerr, N. L. (1978b) Severity of prescribed penalty and mock jurors'

verdicts. Journal of Personality and Social Psychology, 36,

1431-1442.

Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R. W., &

Davis, J. H. (1976) Guilt beyond a reasonable doubt: Effects of

concept definition and assigned decision rule on the judgments of

mock jurors. Journal of Personality and Social Psychology,

34, 282-294.

Kerr, N. L., Bull, R., MacCoun, R. J., & Rathborn, H. (1984) Victim,

culture, and verdict: Modeling juror decision-making. British

Journal of Social Psychology, In press.

Kerr, N. L., & MacCoun, R. J. (1983) Pretrial publicity and juror

judgment: A review of empirical research. Technical report:

Pretrial Publicity Project, American Judicature Society.

Kerr, N. L., & MacCoun, R. J. (1984) The effects of jury size and

polling method on the process and product of jury deliberation.

Under editorial review, Journal of Personality and Social

Psychology.

Knoke, D., & Burke, P. J. (1980) Log-linear models. Sage University

Paper series on Quantitative Applications in the Social Sciences,

07-001. Beverly Hills and London: Sage Publications.

Konecni, V. J., & Ebbesen, E. B. (1982) Social psychology and the law: The choice of research problems, settings, and methodology. In V. J. Konecni & E. B. Ebbesen (Eds.), The criminal justice system: A social-psychological analysis. San Francisco: W. H. Freedman and Co.

Loftus, E. (1983) Whose shadow is crooked? American Psychologist, 38, 576-577.

Marshall, C. R., & Wise, J. A. (1975) 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.

Michelini, R. L., & Snodgrass, S. R. (1980) Defendant characteristics and juridic decisions. Journal of Research in Personality, 14, 340-350.

Mitchell, H. E., & Bryne, D. (1973) The defendant's dilemma: Effects of juror's attitudes and authoritarianism on judicial decisions. Journal of Personality and Social Psychology, 25, 123-129.

Mook, D. G. (1983) In defense of external invalidity. American Psychologist, 83, 379-385.

Myers, D. G., & Kaplan, M. F. (1976) Group-induced polarization in simulated juries. Personality and Social Psychology Bulletin, 2, 63-66.

Myers, D. G., & Lamm, H. (1976) The group polarization phenomenon. Psychological Bulletin, 83, 602-627.

Nagel, S. (1979) Bringing the values of jurors in line with the law. Judicature, 63, 189-195.

Nagel, S., Lamm, D., & Neef, M. (1981) Decision theory and juror decision-making. In B. D. Sales (Ed.), Perspectives in law and psychology. Volume 2: The trial process, (pp.353-386). New York: Plenum Press.

Nagel, S., & Neef, M. (1979) Decision theory and the legal process. Lexington, Mass.: Lexington Books.

Nemeth, C. (1977) Interactions between jurors as a function of majority vs. unanimity decision rules. Journal of Applied Social Psychology, 7, 38-56.

Nisbett, R. E., & Ross, L. (1980) Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Ostrom, T. M., Werner, C., & Saks, M. J. (1978) An integration theory analysis of jurors' presumptions of guilt or innocence. Journal of Personality and Social Psychology, 36, 436-450.

Pennington, N., & Hastie, R. (1981) Juror decision-making models: The generalization gap. Psychological Bulletin, 89, 246-287.

Penrod, S., & Hastie, R. (1979) Models of jury decision-making: A critical review. Psychological Bulletin, 86, 462-492.

Reid, A. H. (1960a) The law of instructions to juries, Vol. 4 (Civil), 3rd Edition. New York: Bobbs-Merrill Co., Inc.

Reid, A. H. (1960a) The law of instructions to juries, Vol. 5
     (Criminal), 3rd Edition. New York: Bobbs-Merrill Co., Inc.

Shaffer, D. R., Case, T., & Brannen, L. (1979) Effects of withheld
     evidence on juridic decisions: Amount of evidence withheld and
     its relevance to the case. Representative Research in Social
     Psychology, 10, 2-15.

Sigall, H., & Ostrove, N. (1975) Beautiful but dangerous: Effects of
     offender attractiveness and nature of crime on juridic judgments.
     Journal of Personality and Social Psychology, 31, 410-414.

Simon, R. J. (1970) "Beyond a reasonable doubt" - An experimental
     attempt at quantification. The Journal of Applied Behavioral
     Science, 6, 203-209.

Simon, R. J., & Mahan, L. (1971) Quantifying burdens of proof: A view
     from the bench, the jury, and the classroom. Law and Society
     Review, 5, 319-330.

Stasser, G, Kerr, N. L., & Bray, R. M. (1982) The social psychology of
     jury deliberations. In N. L. Kerr & R. M. Bray (Eds.), The
     psychology of the courtroom (pp. 221-256). New York: Academic
     Press.

Stasser, G., Kerr, N. L., & Davis, J. H. (1980) Influence processes in
     decision-making groups: A modeling approach. In P. Paulus (Ed.),
     The psychology of group influence. Hillsdale, New Jersey:
     Erlbaum Associates, Inc.

Sue, S., Smith, R. E., & Caldwell, C. (1973) Effects of inadmissable evidence on the decisions of simulated jurors: A moral dilemma. Journal of Applied Social Psychology, 3, 345-353.

Sue, S., Smith, R. E., & Gilbert, R. (1974) Biasing effects of pretrial publicity on judicial decisions. Journal of Criminal Justice, 2, 163-171.

Thomas, E. A. C., & Hogue, A. (1976) Apparent weight of evidence, decision criteria, and confidence ratings in juror decision-making. Psychological Review, 83, 442-465.

Thornton, B. (1977) Effect of rape victim's attractiveness in a jury simulation. Personality and Social Psychology Bulletin, 3, 666-669.

Trial, (1983) Numerical gauge for reasonable doubt ruled prejudicial error. September, 1983, p. 10.

Wolf, S., & Montgomery, D. A. (1977) Effects of inadmissable evidence and level of judicial admonishment to disregard on the judgments of mock jurors. Journal of Applied Social Psychology, 3, 205-219.

[1]
Since the present study uses a variation of Kerr's (1978a) stimulus materials, the victim was portrayed as taking those same precautions.

[2]
Nagel's (e.g., 1979) simplifying assumptions that |UAI| = |UCI|, and that |UAG| = |UCG| were supported in the present study; with t(299) = 1.26, p =.20, and t(300) = 0.84, p = .40, respectively. In fact, these absolute utilities were highly correlated; r = .88, p < .001, and r = .83, p < .001, respectively.

[3]
The Intelligence, Trustworthiness, and Felon scales were basically used as "tie-breakers"; i.e., when there were more than two photographs of a given sex of the same approximate attractiveness level, the photos most neutral on these scales were chosen. Table 2 shows that this attempt was only moderately successful, and more so for the female photos than the male photos. The "what is beautiful is good" stereotype (Berscheid & Walster, 1974) is so robust that it is difficult to manipulate physical attractiveness without manipulating general positivity.

[4]
Log-linear analyses were conducted using the method of partial association (e.g., Brown, 1981). This procedure is analogous to a hierarchical analysis of variance, in that it provides test statistics for each 2-way, 3-way, ...N-way interaction effect. This

is accomplished by fitting a baseline to the data, and then removing

each effect of interest and observing the subsequent decline in

predictive accuracy. For example, in order to test a Verdict x

Instructions effect, a baseline of all possible 2-way effects is fit

to the data, and the likelihood ratio, G^2 , is computed. Next, the

Verdict x Instructions effect is removed and the likelihood ratio is

re-calculated. This latter test statistic is subtracted from the

baseline likelihood ratio, and the resultant G^2 , a test statistic

distributed as χ^2 , is tested against the null hypothesis of

statistical independence; i.e, any differences as a result of dropping

the effect of interest are due to chance. This method is in contrast

to a goodness-of-fit strategy (e.g., Fienberg, 1970) in which the null

hypothesis is that deviations from the hypothesized model are due to

chance.

APPENDIX A

Experimental Materials

MICHIGAN STATE UNIVERSITY

Department of Psychology

DEPARTMENTAL RESEARCH CONSENT FORM

1. I have freely consented to take part in a scientific study being conducted by Robert J. MacCoun under the supervision of Dr. Norbert L. Kerr, Associate Professor of Psychology, MSU.

2. The study has been explained to me and I understand the explanation that has been given and what my participation will involve.

3. I understand that I am free to discontinue my participation in the study at any time without penalty.

4. I understand that the results of the study will be treated in strict confidence and that I will remain anonymous. Within these restrictions, results of the study will be made available to me at my request.

5. I understand that my participation in the study does not guarantee any beneficial results to me.

6. I understand that, at my request, I can receive additional explanation of the study after my participation is completed.

Signed:_____

Title of Experiment: "THE JURY STUDY"

Date:_____

NOTE: You will be asked to read a brief transcript of a criminal trial, and will be asked to respond as you would if you were an actual juror, by completing a questionnaire and deliberating as a group. At the conclusion of the experiment, you may have a number of questions about the research. If so, you are invited to attend a discussion session conducted by the experimenter on Friday, March 30th at 1 pm at 412 Baker Hall, or you may call Rob MacCoun at 353-6611.

THIS EXPERIMENT WILL LAST NO LONGER THAN ONE AND A HALF HOURS AND YOU WILL RECEIVE THREE EXPERIMENTAL RESEARCH CREDITS FOR YOUR PARTICIPATION.

117

THE JURY STUDY

Your initials:
Your sex:
Date:

1. In my opinion, the defendant, William Lambeth, is:

   _____ Guilty of auto theft

   _____ Not Guilty of auto theft

2. How confident are you in the above verdict?
          (circle one number)

   No   Complete
   confidence 0 1 2 3 4 5 6 7 8 9 10 confidence

3. In my opinion, the probability that William Lambeth did commit the charged offense is (place a check mark somewhere along the following scale):

   :_____:_____:
   0   .5   1.0

   it's certain   there's a   it's certain
   that Lambeth   50-50 chance   that Lambeth
   did NOT steal   that Lambeth   DID steal the
   the car   stole the car   car

4. What is the smallest probability of guilt that you believe would be necessary in order to conclude that Lambeth is GUILTY of auto theft (i.e., if there were any less than that probability, you'd vote NOT GUILTY)?

   :_____:_____:
   0   .5   1.0

   No evidence   Complete evidence
   of guilt   of guilt

5. Suppose for a minute that William Lambeth were found guilty, and that you were the judge who had to sentence him. Assuming that the maximum penalty is 20 years imprisonment, what prison sentence would you recommend?

          _____ years, _____ months

The fact is that William Lambeth either did or did not steal Helen Bednard's car. Furthermore, there are two possible decisions that can be made at the end of the trial: Lambeth can be found GUILTY, or Lambeth can be found NOT GUILTY. Thus, there are four possible outcomes of William Lambeth's trial:

TRUE STATE OF THE WORLD

	Lambeth DID steal the car	Lambeth DID NOT steal the car
VERDICT		
Lambeth is found GUILTY	A guilty man is convicted	An innocent man is convicted
Lambeth is found NOT GUILTY	A guilty man is set free	An innocent man is set free

6. Please consider each possible outcome of the trial and indicate whether you feel that the outcome is a POSTIVE outcome or a NEGATIVE outcome:

TRUE STATE OF THE WORLD

	Lambeth DID steal the car	Lambeth DID NOT steal the car
VERDICT		
Lambeth is found GUILTY	___ positive ___ negative	___ positive ___ negative
Lambeth is found NOT GUILTY	___ positive ___ negative	___ positive ___ negative

7. Now consider each possible outcome of the trial and write a number in each square to indicate HOW postive or HOW negative that outcome would be if it occurred. If you think that an outcome would be positive, use any number between zero and positive infinity. If you think an outcome would be negative, use any number between zero and negative infinity.

TRUE STATE OF THE WORLD

	Lambeth DID steal the car	Lambeth DID NOT steal the car
VERDICT		
Lambeth is found GUILTY		
Lambeth is found NOT GUILTY		

8. In my opinion, the probability that William Lambeth did commit the charged offense is (check one):

_____ 0 chances in 10 (i.e.,it's certain that Lambeth did NOT
                       steal the car)
_____ 1 chance in 10

_____ 2 chances in 10

_____ 3 chances in 10

_____ 4 chances in 10

_____ 5 chances in 10 (i.e., there's a 50-50 chance that Lambeth
                       stole the car)
_____ 6 chances in 10

_____ 7 chances in 10

_____ 8 chances in 10

_____ 9 chances in 10

_____ 10 chances in 10 (i.e., it's certain thsat Lambeth DID
                        steal the car)

9. What is the smallest probability of guilt that you believe would be necessary in order to conclude that Lambeth is GUILTY of auto theft (i.e., if there were any less than that probability, you'd vote NOT GUILTY)?

_____ 0 chances in 10 (i.e., no evidence of guilt)

_____ 1 chance in 10

_____ 2 chances in 10

_____ 3 chances in 10

_____ 4 chances in 10

_____ 5 chances in 10

_____ 6 chances in 10

_____ 7 chances in 10

_____ 8 chances in 10

_____ 9 chances in 10

_____ 10 chances in 10 (i.e., complete evidence of guilt)

10. Please rate your impressions of William Lambeth, the defendant, by placing a check mark on each of the following scales:

Believable   :____:____:____:____:____:____:____: Unbelievable

Likeable   :____:____:____:____:____:____:____: Not Likeable

Attractive   :____:____:____:____:____:____:____: Unattractive

Intelligent :____:____:____:____:____:____:____: Unintelligent

11. Please rate your impressions of Helen Bednard, the victim, by placing a check mark on each of the following scales:

Believable   :____:____:____:____:____:____:____: Unbelievable

Likeable   :____:____:____:____:____:____:____: Not Likeable

Attractive   :____:____:____:____:____:____:____: Unattractive

Intelligent :____:____:____:____:____:____:____: Unintelligent

12. Please indicate how important the judge's instructions were in helping you to determine whether or not there was sufficient evidence in the trial to convict William Lambeth: (circle one)

    1   2   3   4   5   6   7

Completely   Completely
unimportant   important

13. Please indicate how comprehensible the judge's instructions were for you: (circle one)

    1   2   3   4   5   6   7

Completely   Completely
incomprehensible   comprehensible

14. Please complete the following sentence by writing a number in the blank:

"It is better to let _____ guilty defendant(s) go free than to convict one innocent defendant."

15. How sympathetic did you feel towards William Lambeth, the defendant (circle one)?

    1   2   3   4   5   6   7

Very   Very
unsympathetic   sympathetic

16. How sympathetic did you feel towards Helen Bednard, the victim (circle one)?

           1   2   3   4   5   6   7

Very   Very
unsympathetic   sympathetic

17. In my opinion, the odds that William Lambeth did commit the charged offense are (write a number in each blank):

           _____ to _____

18. What are the smallest odds of guilt that you believe would be necessary in order to conclude that William Lambeth is GUILTY of auto theft (i.e., if the odds were any smaller than that, you'd vote NOT GUILTY)?

           _____ to _____

FOREPERSON'S QUESTIONNAIRE

After each of the members of your group have finished filling out the
individual questionnaires, you should close your door and begin
deliberation. Please discuss the case as a group and attempt to reach
a unanimous group verdict. (Note that there is a microphone in
your rooom. When you close the door, the experimenter will begin tape-
recording your deliberation. You may find that once you begin discussing
the case, the presence of the microphone will be easy to ignore). Your
group may deliberate until _____. Your experimenter will
notify you when you have only 5 minutes left to deliberate.

Foreperson's initials:_____

Date:_____

Time your jury began deliberation:_____

Time your jury completed deliberation (i.e., reached a unanimous group
    verdict or "hung"): _____

GROUP VERDICT (at completion of deliberation):

    "We find William Lambeth _____ of the charge of auto theft"
                        (check one)

                    _____ GUILTY

                    _____ NOT GUILTY

                    _____ HUNG (i.e., we were unable to reach a
                                unanimous group verdict in the
                                time allotted)

THE JURY STUDY

Your initials:
Your sex:
Date:

Jurors, please fill this short questionnaire out AFTER your jury
has deliberated and reached a unanimous group verdict.

1. What is your personal verdict? _____ GUILTY _____ NOT GUILTY

2. How confident are you in the above verdict? (circle one number)

   NO CONFIDENCE 0 1 2 3 4 5 6 7 8 9 10 COMPLETE CONFIDENCE

3. How satisfied are you with your group's verdict? (circle one number)

   VERY   VERY
   DISSATISFIED 0 1 2 3 4 5 6 7 8 9 10 SATISFIED

4. In my opinion, the probability that William Lambeth did commit the
   charged offense is: (fill in the blank with a number from 0 to 10)

        "_____ chances in 10"

5. In my opinion, the odds that William Lambeth did commit the charged
   offense are (write a number in each blank):

        "_____ to _____"

6. In my opinion, the probability that William Lambeth did commit the
   charged offense is (place a check mark somewhere along the following
   scale):

        :_____:_____:
        0   .5   1.0

        it's certain   there's a   it's certain
        that Lambeth   50-50 chance   that Lambeth
        did NOT steal   that Lambeth   DID steal the
        the car   stole the car   car

We'd appreciate any comments you'd like to make about this study. Did
you enjoy it? If so, why? If not, why not? Was there anything you
found confusing or hard to understand?

        Thank you for your interest and participation...

APPENDIX B

Analysis of Variance and

Log-Linear Modeling Tables

Table B-1

Analysis of Variance: Victim Attractiveness

Manipulation Check by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	.41	.24
Instructions	1	.19	.11
Victim Att.	1	375.20	219.63 **
Defend. Att.	1	.59	.35
S x I	1	.27	.16
S x V	1	2.70	1.58
S x D	1	.03	.02
I x V	1	1.09	.64
I x D	1	.56	.33
V x D	1	.03	.01
S x I x V	1	14.08	8.24 *
S x I x D	1	.00	.00
S x V x D	1	.42	.25
I x V x D	1	.26	.15
S x I x V x D	1	.76	.44
Error	301	1.71	

* p < .01
** p < .001

Table B-2

Analysis of Variance: Defendant Attractiveness

Manipulation Check by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	5.30	4.71 *
Instructions	1	.98	.87
Victim Att.	1	.63	.56
Defend. Att.	1	341.55	303.74 ***
S x I	1	1.44	1.28
S x V	1	.47	.42
S x D	1	.78	.69
I x V	1	.04	.03
I x D	1	1.14	1.02
V x D	1	.96	.85
S x I x V	1	.10	.09
S x I x D	1	15.43	13.72 **
S x V x D	1	1.03	.92
I x V x D	1	.69	.62
S x I x V x D	1	1.49	1.32
Error	302	1.12	

* p < .05
** p < .001

Table B-3

Log-Linear Analysis: Individual Pre-Deliberation
Verdicts by Subject Sex, Instructions,
Victim and Defendant Attractiveness

Effect	df	G^2
Baseline	16	15.65
Verdict x Sex	1	.06
Verdict x Instructions	1	7.10 *
Verdict x Victim Att.	1	.00
Verdict x Defendant Att.	1	.51
Baseline	7	-9.43
Verdict x S x I	1	.33
Verdict x S x V	1	3.37
Verdict x S x D	1	.04
Verdict x I x V	1	1.56
Verdict x I x D	1	2.99
Verdict x V x D	1	.19
Baseline	1	-5.27
Verdict x S x I x V	1	.12
Verdict x S x I x D	1	.13
Verdict x S x V x D	1	3.88
Verdict x I x V x D	1	.00

* $p < .05$

Table B-4

Analysis of Variance: Pre-Deliberation

Guilt Scores by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	13.90	.22
Instructions	1	453.31	7.27 ***
Victim Att.	1	.01	.00
Defend. Att.	1	9.91	.16
S x I	1	25.75	.41
S x V	1	189.91	3.05
S x D	1	.52	.01
I x V	1	36.12	.58
I x D	1	136.01	2.18
V x D	1	.93	.02
S x I x V	1	3.70	.06
S x I x D	1	10.15	.16
S x V x D	1	163.67	2.63
I x V x D	1	2.10	.03
S x I x V x D	1	37.72	.61
Error	305	62.33	

* p < .01

Table B-5

Analysis of Variance: Pre-Deliberation

Guilt Score Internal Analysis by Subject Sex,

Instructions, Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	33.06	.55
Instructions	1	169.65	2.83
Victim Att.	1	2.81	.05
Defend. Att.	1	39.79	.66
S x I	1	.03	.00
S x V	1	292.03	4.86 *
S x D	1	.80	.01
I x V	1	10.44	.17
I x D	1	254.44	4.24 *
V x D	1	50.74	.85
S x I x V	1	.50	.01
S x I x D	1	21.92	.37
S x V x D	1	45.03	.75
I x V x D	1	22.07	.37
S x I x V x D	1	82.88	1.38
Error	177	60.06	

* p < .05

Table B-6

Analysis of Variance: Recommended Sentences

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	412.45	.14
Instructions	1	11745.94	3.85
Victim Att.	1	270.56	.09
Defend. Att.	1	86.15	.03
S x I	1	943.36	.31
S x V	1	1962.57	.64
S x D	1	244.51	.08
I x V	1	9298.06	3.05
I x D	1	478.31	.16
V x D	1	623.63	.21
S x I x V	1	50.91	.02
S x I x D	1	1328.99	.44
S x V x D	1	139.58	.05
I x V x D	1	1.56	.00
S x I x V x D	1	536.07	.18
Error	286	3047.76	.62

Table B-7

Analysis of Variance: Victim Believability

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	.64	.38
Instructions	1	5.21	3.14
Victim Att.	1	2.57	1.55
Defend. Att.	1	.01	.01
S x I	1	8.77	5.28 *
S x V	1	.34	.20
S x D	1	.21	.13
I x V	1	2.86	1.72
I x D	1	3.12	1.88
V x D	1	.18	.11
S x I x V	1	.00	.00
S x I x D	1	.34	.21
S x V x D	1	.99	.60
I x V x D	1	.03	.02
S x I x V x D	1	3.28	1.98
Error	301	1.66	

* $p < .05$

Table B-8

Analysis of Variance: Victim Likeability

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	.84	.54
Instructions	1	1.29	.83
Victim Att.	1	25.13	16.16 *
Defend. Att.	1	1.77	1.14
S x I	1	.91	.59
S x V	1	.10	.07
S x D	1	1.74	1.21
I x V	1	2.79	1.79
I x D	1	.04	.03
V x D	1	.08	.05
S x I x V	1	.09	.06
S x I x D	1	.00	.00
S x V x D	1	1.00	.64
I x V x D	1	.49	.32
S x I x V x D	1	.14	.09
Error	301	1.56	

* p < .001

Table B-9

Analysis of Variance: Victim Intelligence

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	.00	.00
Instructions	1	.70	.55
Victim Att.	1	5.44	4.24 *
Defend. Att.	1	.35	.27
S x I	1	.42	.33
S x V	1	.00	.00
S x D	1	.00	.00
I x V	1	1.90	1.48
I x D	1	.13	.10
V x D	1	.09	.07
S x I x V	1	5.58	4.34
S x I x D	1	.90	.70
S x V x D	1	2.48	1.93
I x V x D	1	.80	.62
S x I x V x D	1	.04	.03
Error	301	1.28	

* p < .05

134

Table B-10

Analysis of Variance: Sympathy for Victim

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	6.22	2.64
Instructions	1	2.95	1.25
Victim Att.	1	1.46	.62
Defend. Att.	1	.41	.18
S x I	1	.40	.17
S x V	1	.44	.19
S x D	1	.91	.38
I x V	1	2.89	1.23
I x D	1	1.10	.47
V x D	1	1.27	.54
S x I x V	1	4.42	1.03
S x I x D	1	.00	.00
S x V x D	1	.11	.05
I x V x D	1	4.58	1.95
S x I x V x D	1	20.59	8.74 *
Error	299	2.36	

* p < .01

Table B-11

Analysis of Variance: Defendant Believability

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	.88	.36
Instructions	1	13.26	5.44 *
Victim Att.	1	2.85	1.17
Defend. Att.	1	.74	.30
S x I	1	4.35	1.79
S x V	1	1.17	.49
S x D	1	.34	.14
I x V	1	.21	.09
I x D	1	.12	.05
V x D	1	.01	.01
S x I x V	1	1.60	.66
S x I x D	1	.07	.03
S x V x D	1	1.37	.56
I x V x D	1	.34	.14
S x I x V x D	1	.03	.01
Error	302	2.44	

* p < .05

136

Table B-12

Analysis of Variance: Defendant Likeability

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	.21	.15
Instructions	1	.23	.16
Victim Att.	1	1.31	.90
Defend. Att.	1	6.71	4.62 *
S x I	1	.61	.42
S x V	1	.06	.04
S x D	1	1.35	.93
I x V	1	2.65	1.83
I x D	1	1.14	.79
V x D	1	2.09	1.44
S x I x V	1	.02	.02
S x I x D	1	4.63	3.19
S x V x D	1	1.75	1.21
I x V x D	1	1.05	.72
S x I x V x D	1	1.35	.93
Error	302	1.45	

* p < .05

Table B-13

Analysis of Variance: Defendant Intelligence

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	2.23	1.86
Instructions	1	.94	.78
Victim Att.	1	.82	.68
Defend. Att.	1	11.03	9.19 **
S x I	1	2.55	2.13
S x V	1	1.23	1.03
S x D	1	.11	.09
I x V	1	.03	.03
I x D	1	.22	.18
V x D	1	6.88	5.74 *
S x I x V	1	.42	.35
S x I x D	1	.17	.14
S x V x D	1	.17	.14
I x V x D	1	.07	.06
S x I x V x D	1	2.96	2.47
Error	302	1.20	

* p < .05
** p < .01

Table B-14

Analysis of Variance: Sympathy for Defendant

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Subject Sex	1	.20	.11
Instructions	1	1.28	.71
Victim Att.	1	.11	.06
Defend. Att.	1	3.18	1.76
S x I	1	.45	.25
S x V	1	9.25	5.11 *
S x D	1	5.98	3.30
I x V	1	.41	.23
I x D	1	.35	.19
V x D	1	.19	.11
S x I x V	1	1.76	.97
S x I x D	1	.88	.49
S x V x D	1	.39	.22
I x V x D	1	.27	.15
S x I x V x D	1	18.82	10.40 **
Error	299	1.81	

 * $p < .05$
** $p < .001$

Table B-15

Log-Linear Analysis: Group Verdicts

by Subject Sex, Instructions,

Victim and Defendant Attractiveness

Effect	df	G^2
Baseline	27	26.40
Verdict x Size	2	7.88 *
Verdict x Instructions	2	.57
Verdict x Victim Att.	2	3.45
Verdict x Defendant Att.	2	7.25 *
Baseline	11	-17.62
Verdict x S x I	2	4.85
Verdict x S x V	2	.30
Verdict x S x D	2	2.92
Verdict x I x V	2	.71
Verdict x I x D	2	5.52
Verdict x V x D	2	.82

* p < .01

Table B-16

Repeated Measures ANOVA: Guilt Scores by
Time, Size, Instructions, and
Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Size	1	150.71	2.84
Instructions	1	52.74	0.99
Victim Att.	1	18.08	0.34
Defend. Att.	1	14.89	0.28
S x I	1	127.70	2.41
S x V	1	4.11	0.08
S x D	1	79.76	1.50
I x V	1	83.77	1.58
I x D	1	14.93	0.28
V x D	1	0.31	0.01
S x I x V	1	49.87	0.94
S x I x D	1	3.64	0.07
S x V x D	1	7.53	0.14
I x V x D	1	6.84	0.13
S x I x V x D	1	11.50	0.22
Error	71	53.07	

Table B-16 (Continued)

Repeated Measures ANOVA: Guilt Scores by

Time, Size, Instructions, and

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Time	1	106.88	14.44 **
T x Size	1	3.94	.53
T x Instructions	1	.04	.01
T x Victim Att.	1	8.48	1.15
T x Defend. Att.	1	55.36	7.48 *
T x S x I	1	.37	.05
T x S x V	1	6.59	.89
T x S x D	1	26.70	3.61
T x I x V	1	4.20	.57
T x I x D	1	4.19	.57
T x V x D	1	8.85	1.20
T x S x I x V	1	9.29	1.25
T x S x I x D	1	4.36	.59
T x S x V x D	1	6.89	.93
T x I x V x D	1	22.28	3.01
T x S x I x V x D	1	3.26	.44
Error	71	7.40	

* $p < .01$
** $p < .001$

Table B-17

Repeated Measures ANOVA: Mean p(G) Estimates

by Time, Size, Instructions, and

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Size	1	445.39	0.97
Instructions	1	44.88	0.10
Victim Att.	1	145.09	0.32
Defend. Att.	1	1.91	0.00
S x I	1	324.66	0.71
S x V	1	271.68	0.59
S x D	1	497.53	1.08
I x V	1	236.76	0.52
I x D	1	42.09	0.09
V x D	1	68.66	0.15
S x I x V	1	0.02	0.00
S x I x D	1	76.18	0.17
S x V x D	1	111.17	0.24
I x V x D	1	14.63	0.03
S x I x V x D	1	438.98	0.96
Error	68	458.68	

Table B-17 (Continued)

Repeated Measures ANOVA: Mean p(G) Estimates

by Time, Size, Instructions, and

Victim and Defendant Attractiveness

Source	df	Mean Square	F-Ratio
Time	1	440.52	5.78 *
T x Size	1	6.11	.08
T x Instructions	1	245.17	3.21
T x Victim Att.	1	12.54	.16
T x Defend. Att.	1	240.78	3.16
T x S x I	1	383.45	5.03 *
T x S x V	1	10.49	.14
T x S x D	1	28.51	.37
T x I x V	1	111.06	1.46
T x I x D	1	123.47	1.62
T x V x D	1	7.44	.10
T x S x I x V	1	74.84	.98
T x S x I x D	1	.07	.00
T x S x V x D	1	28.17	.37
T x I x V x D	1	91.48	1.20
T x S x I x V x D	1	93.09	1.22
Error	68	76.27	

* p < .05