105015

86-IJ-CX-0004
FINAL REPORT

PROBABILISTIC SCIENTIFIC EVIDENCE: JURORS' INFERENCES

by

Jane Goodman GRF

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

1986

Approved by: Donald J. Mizokawa
Chairperson of Supervisory Committee
Department of Educational Psychology

Approved by: Elizabeth F. Loftus
Dissertation Adviser
Department of Psychology

August, 1986

University of Washington

Abstract

PROBABILISTIC SCIENTIFIC EVIDENCE: JURORS' INFERENCES

by Jane Goodman

Chairperson of the Supervisory Committee:

Donald T. Mizokawa, Associate Professor

Department of Educational Psychology

Dissertation Adviser:

Elizabeth F. Loftus, Professor

Department of Psychology

Two pencil-and-paper jury simulation studies investigated inferences drawn by mock-jurors from probabilistic frequency evidence. Study One varied the frequency of a suspect blood type as either 10%, 5%, 1%, .1% or no frequency evidence was provided. Subjects were 233 psychology undergraduates who returned verdicts individually. More weight was accorded to the probabilistic evidence than to other facts, but subjects failed to distinguish between frequency probabilities 5%, 1%, and .1%. Comparisons with Bayesian "rational" estimates showed that subjects in all groups underused the probability evidence. One-way analyses of variance yielded significant differences in group estimates of guilt, doubt of guilt, and the extent to which the frequency evidence and matching blood-type evidence were accorded incriminating weight. No support for the prosecutor's fallacy was found, i.e., subjects did not confuse the burden of proof and the probability evidence. Error rates on questions about the probabilistic evidence were high.

Study Two was a 2x2x2 design. Subjects were 223 jurors who read summaries of an arson case involving circumstantial evidence based on forensic analyses of gasoline samples. The first variable, burden of proof, had two levels: the civil standard (preponderance of the evidence) and the criminal standard (beyond a reasonable doubt). The second variable was the linguistic form of the probabilistic testimony, expressed as odds (1 in 1000) and as a percentage (.1%). The third variable was the presence or absence of a visual aid depicting results of chromatographic tests.

Burden of proof had no effect on the proportion of findings of arson or estimates of defendant's guilt. Neither the linguistic form of the probabilistic evidence nor the presence or absence of visual aid had a significant effect on mock-jurors' verdicts. Significantly more weight was accorded to the expert witness who presented scientific evidence than to the expert who did not. No support was found for the hypothesis that jurors are susceptible to the prosecutor's fallacy. Error rates on questions about the probabilistic evidence were high. Some deficiencies in mock-jurors' abilities to consider and weigh expert testimony were found; however, they indicate that there is less danger that jurors will overbelieve or overuse scientific, quantitative evidence than some judges have feared.

TABLE OF CONTENTS

LIST OF TABLES

Number | Page

LIST OF FIGURES

ACKNOWLEDGEMENTS

This project is based on work supported by Grant # 86-IJ-CX-0004 from the National Institute of Justice, U.S. Department of Justice and Grant # SES-8600638 from the National Science Foundation. Opinions, findings, conclusions or recommendations stated in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice or the views of the National Science Foundation. Additional funds were provided by a Grant-in-Aid from Division 41, American Psychological Association.

I thank my adviser and good friend, Elizabeth Loftus, for the many things she has taught me in connection with this project and others. I thank Donald Mizokawa for his encouragement and guidance, and especially for the enjoyable conversations along the way. Edie Greene and Larry Severance have been inspiring wayfarers in law and psychology. I am grateful to Robert Abbott, Earl Butterfield, Timothy Standal and Moncrieff Smith for their critique and support.

A number of others deserve credit. George Ishii and Dale Mann of the King County Crime Laboratory aided me in preparing up-to-date, realistic trial materials. Robert Cannon, Judy Rutledge and Nicola Hommel at the King County Courthouse permitted me to gather data from jurors at the Courthouse. Karsten Moen, Vicki Lai, and Jonathan Fraser assisted with data collection and analysis. Finally, I thank my husband, Oscar, whose patience and enthusiasm exceeded all my probability estimates.

INTRODUCTION

The Nature of Legal Evidence

Evidence presented in the course of a trial to prove a case is in one of two forms: direct or circumstantial. Direct evidence, if true, establishes a fact itself. For example, direct evidence includes testimony by witnesses who describe what they saw, heard, tasted, smelled or felt, based on their personal observations and knowledge. According to Braun (1982), in general, the role of the trier of fact is simply to "judge the credibility of the witnesses in ascertaining the truth or existence of the principal fact." Circumstantial evidence, by comparison, is proof of facts and circumstances which permit a reasonable inference that other events took place. Thus, circumstantial evidence proves a fact indirectly, requiring the jury to infer the truth or existence of the principal fact given the proven facts.

Braun (1982) noted that in cases involving circumstantial evidence, there is always the possibility that the inference drawn will be erroneous. Of course, direct testimony can also be erroneous (Loftus, 1980), although many courts assume that the opportunity to cross-examine a witness who provides direct testimony provides an adequate safeguard of the veracity of the evidence (Abney, 1986). This safeguard is absent in the case of circumstantial evidence (Sperlich, 1985; Loh, 1985). Scientific evidence may be either direct or circumstantial, but is more typically the latter.

The Increasing Use of Scientific Evidence

Curtis and Wilson (1979) noted the dramatic increase in the use of scientific evidence in litigation both in civil and in criminal cases. In a survey of judges and attorneys (National Center for State Courts, 1980), the respondents reported that scientific evidence was admitted in approximately one third of their trials. According to Imwinkelried (1983) the increasing use of this evidence is partially attributable to the pace of technological change; partially to a more liberalized application in the evidentiary barriers to the admissibility of scientific proof, and partially to jurors' expectations nowadays that scientific proof will be offered.

Prior to 1980, statistical evidence was something of a novelty. In a thorough review of the uses of statistics and social science in litigation, Loh (1979) noted that the major use of statistics was in the adjudication of quantitative factual issues, such as making inferences about the numerical characteristics of a population based on sampling operations, e.g., in civil anti-trust cases. The use of statistics to adjudicate the likelihood of a particular event was rare and was approached with considerable caution, particularly in criminal cases.

In the past decade, however, the use of statistics to assess the likelihood of a particular event has become more commonplace, in both civil and criminal trials (Panel on Statistical Assessments as Evidence in the Courts, 1985). In civil suits, scientific evidence is often introduced to prove causation in tort cases such as products liability cases or medical malpractice suits. Statistical evidence typically

comprises the basis of proof of intentional discrimination in housing, employment, voting and other civil rights cases. Giannelli (1983) noted that since the 1970s, more criminal cases have been prosecuted on the basis of circumstantial evidence--much of which is novel scientific evidence such as sound spectometry, neutron activation analysis, gun-shot residue tests, ion microprobic analysis, trace metal detection, psycholinguistics, fingernail comparisons, enzyme blood testing, gas chromatographic analysis, battered wife syndrome evidence, rape trauma syndrome evidence, and hypnotically-refreshed testimony. According to one prosecutor (Clark, 1969), "the backbone of every circumstantial evidence case" (p. 369) is scientific proof.

The Form of Scientific Evidence: Expert Testimony

Scientific evidence is introduced through an expert witness who generally describes some of the scientific methods used in conducting a study of a particular topic, and who then provides some numerical or statistical information based on the instrumental procedures used, and/or the results of the study, such as incidence rates of a set of characteristics in a given sample. The expert may also make a statement about the probability or likelihood of finding a certain characteristic in a sample by chance alone, i.e., a conditional probability statement. The purpose of entering probability statements into evidence is to ascribe weight to the circumstantial, scientific evidence.

As the use of scientific evidence has increased, controversy over the form of this expert testimony has increased concomitantly (Imwinkelried, 1981). Not only do standards regarding what the expert

may or may not say vary enormously from one courtroom to the next, but the question of juror competency to understand and apply the expert testimony has been raised (Tribe, 1971).

The Controversy: Scientific Evidence in Jury Trials

Numerous theories of juror responses to probabilistic testimony have been advanced by judges, lawyers and psychologists, culminating in hypotheses that jurors tend either to overweight, underweight or ignore this evidence. First, as Loh (1979) suggested, jurors may be mesmerized by the apparent precision of mathematical evidence, and therefore accord it undue weight. Second, because of their lack of training and expertise with the subject matter, jurors are believed simply to misconstrue the evidence (Austin, 1984). For example, misunderstanding of quantitative information is thought to result in confusion between probabilistic testimony and the burden of proof (Thompson, 1986). Finally, jurors are presumed to ignore scientific testimony because it is dull in comparison with anecdotal information (Saks & Kidd, 1980), or because they fail to follow its implications (Burger, 1980).

Judicial recognition of the problem has come from many sources (Thomas, 1983; Tribe, 1971), including former Chief Justice of the United States Supreme Court, Warren Burger. A strong advocate for the abolition of juries in complex cases, he expressed the view that scientific concepts or business disputes place an undue burden on jurors. Consequently, the exclusion of probabilistic scientific evidence from jury trials has been urged. Precisely what makes a case complex is less easy to define. Chief Justice Burger (1980) enumerated

three potential sources of difficulty: (a) comprehension of facts; (b) recall of facts; (c) application of the law to the facts:

> There is a limit to the capacity of any of us--jurors or even judges--to understand and remember a mass of complicated transactions described in a long trial (p. 450).

Burger went on to suggest that when complicated evidence is involved, in lieu of a jury trial, attorneys should stipulate to trial before three judges instead of one judge, to avoid the problem of having the case assigned to a judge less "sophisticated in complex economic, business or environmental cases" (p. 457). However, there is little evidence to suggest that a judge's performance in the face of probabilistic testimony necessarily differs from that of jurors. Comparisons of judge and jury decisions in general have shown considerable agreement between the two groups (Broeder, 1959; Kalven & Zeisel, 1966). Accordingly, other legal theorists and proponents of the jury system favor the continued presentation of this type of evidence to jurors, but advocate the development of specific guidelines regarding its form, to assist jurors in understanding and applying the evidence, and to guard against invading the province of the jury.

The Need for Empirical Resolutions

Despite current preoccupation with the issue, little is known about the way in which jurors perceive and use probabilistic scientific evidence and other complex evidence in reaching a verdict. Preliminary studies have confirmed that jurors do not consider evidence in a Bayesian fashion (Faigman, 1983). What jurors do instead bears further

investigation. The need for systematic empirical research to resolve these questions has been acknowledged by both lawyers (Burger, 1980) and by social scientists (Saks & Kidd, 1980; Lempert, 1981).

In commenting on the misuse and nonuse of science in law, and the ensuing mutual disparagement by lawyers and scientists, Sperlich (1985), a political scientist, attributed some of the interdisciplinary tension to the deductive nature of legal reasoning, which seeks absolute certainty, while the empirical-inductive method of science, by comparison, can only offer probabilities. He noted:

> Few observers are satisfied with the current state of scientific evidence in litigation; many agree that matters should not remain as they are. What is needed, however, are not simply changes, but improvements. The task of **evidentiology** is to identify the changes that help. Their implementation will make our science less dismal. (p. 352). [Emphasis in the original.]

Sperlich did not endorse the viewpoint that the jury system must be abandoned in cases involving scientific evidence. Included in the list of improvements to the system which he advocated is the development of techniques designed to assist jurors. Specifically, he argued that jurors need assistance (1) in distinguishing data, generalizations and opinions which meet scientific standards from those that do not, including detecting non-standard operations, such as surveys with too small a sample, premature generalizations, or experiments without control groups; and (2) in assessing the adequacy or inadequacy of expert witnesses.

Empirical research on the impact of scientific evidence on jurors is in its very early phases. Monahan and Loftus (1983) noted that much

prior jury research has focused on the effects of procedural rules, such as jury instructions or the order in which information is presented. A second major line of inquiry has been research focused on substantive topics such as plea bargaining, judicial attitudes, and the credibility of the communicator (eyewitness, attorney, judge, or defendant). Much past psychological research has been on evidence as evidence: eyewitness testimony, expert testimony, witness examination, and the influence of extra-legal factors such as a defendant who fails to take the stand. By comparison, as noted by Sperlich (1985), studies of the use of scientific evidence by judges and jurors are rare.

The Present Research

The purpose of the present research to gather preliminary information on mock-jurors' understanding of circumstantial, scientific evidence which contains probabilistic information. The goals of the two studies are (1) to refine the task analysis, (2) to advance the theory of mock-jurors' task performance; and, (3) to determine whether mock-jurors can be assisted in drawing appropriate inferences from probabilistic evidence by minimum intervention.

Overview of the Chapters

Chapter I outlines current case law and prevailing legal perspectives on issues raised by the admission of probabilistic evidence in jury trials. Considered judicial opinions on the ability of jurors to make rational decisions in cases in which such evidence is admitted are categorized and summarized. In Chapter II, psychological literature which bears on these judicial hypotheses is reviewed.

Chapters III and IV describe the present research. The first study examines the influence of varying frequency probabilities upon the weight accorded to evidentiary facts and mock-jurors' perceptions of the defendant's guilt. The second study examines the effect of varying the linguistic form of the probability information, of providing a visual aid, and also varies the burden of proof or decision-rule against which the evidence is evaluated. Research conclusions are presented in Chapter V.

CHAPTER ONE

THE LEGAL PERSPECTIVE:

IS PROBABILISTIC SCIENTIFIC EVIDENCE PROPER SUBJECT MATTER FOR A JURY?

Overview of Probabilistic Scientific Evidence

One of the earliest cases in which probabilistic circumstantial played a prominent role (Mode, 1963) was the infamous French trial which began in 1894, accusing Captain Alfred Dreyfus of treason. Because a suspect document purportedly written by Dreyfus contained a distribution of alphabetical characters similar to those in letters known to be written by him, and because the suspect document contained a distribution of characters different from that found in average French prose, the prosecution inferred first, that its author was Dreyfus, and second, that it contained secret codes, proving Dreyfus was a traitor. According to Hedrick (1928), defense testimony by the most famous mathematician of France, Henri Poincaré, to the effect that the most probable frequency distribution was not the most probable character distribution in this case simply baffled the court and did little to exonerate Dreyfus. A second mathematician, Painlevé, who was more familiar with lawyers and court procedures (he later became the French Prime Minister), was more successful in casting doubt on the prosecution's case, when he explained that even Racine, the greatest French tragedian, would be a traitor by this standard, for his writing too did not contain the most probable distribution of letters of the alphabet.

Not all probabilistic evidence generates the scandal and furor which surrounded the Dreyfus case, but controversy is common. Issues posed by the introduction of probabilistic evidence have been reviewed

by Hallock (1977) and Jonakait (1983). An example of a more recent case involved probabilistic, circumstantial evidence regarding the similarity of hair samples and fibers from a green carpet taken from the home of suspect Wayne Williams arrested in connection with a chain of homicides in Atlanta, Georgia. The fiber sample evidence played such a pivotal role in linking Williams to one of the homicide victims, that this was called "the case of the green carpet" (Time, 1981).

Circumstantial scientific evidence routinely introduced at trial nowadays includes "partial transfer evidence" such as fingerprints, blood types (human leukocyte antigen analyses), ballistics, footprint comparisons, and analyses of hair samples, clothing fibers, teeth-marks, and particle evidence (neutron activation analyses of glass, paint, dirt, etc.). In addition to transfer evidence, other types of probabilistic evidence listed by Monahan and Walker (1985) include various medical tests, and an increasing variety of psychological "syndrome" evidence, such as post-rape trauma syndrome, battered wife syndrome, child abuse syndrome, Vietnam veteran syndrome, etc. The evidence may be introduced by either the prosecution/plaintiff or the defense, depending on whether the object of the testimony is to prove that someone is included in a particular class or excluded from a particular class. For instance, blood type evidence may be introduced in a criminal case to show that blood samples from the scene of a crime match those of a suspect, or may be introduced in a civil paternity suit, to exclude a putative father from a possible class of parents.

Wehmhoefer (1985) and Braun (1982) noted that probabilistic expert testimony is based either on a frequency probability model or the

classical probability model of LaPlace. Frequency probability evidence is the most common form of scientific evidence used in criminal cases when direct evidence is lacking to predict what may or may not have occurred. Frequency ratios can only be applied when a set of circumstances has been observed enough times in the past to establish a meaningful proportion. Witnesses called to testify may be experts drawn from a number of disciplines, such as forensic scientists, engineers, psychologists, medical experts, etc., who rely on their experience. By comparison, classical probability proportions are not calculated by appeals to empirical observations made by experts in a particular field, such as fingerprint analysis, or chromatographic analyses, etc., but are calculated solely by the application of mathematical formulas. These probability ratios can be produced without the necessity of relying on experience. Expert witnesses in classical probability cases are therefore generally mathematicians, statisticians, or economists.

Classical probability theory may be applied where the defendant contends that circumstances giving rise to criminal action were brought upon by fortuitous happenings (such as an accident or coincidence), and that he or she acted in a neutral and random manner. The issue is to determine the chance that the circumstances could have arisen from purely random occurrences. If the probability is close to zero, the factfinder can conclude that the circumstances resulted at least in part from the deliberate acts of someone. This is how classical probability theory is employed to determine whether an employer is guilty of employment discrimination. As with frequency probability

models, the probability proportion should be close to zero or one to imply assured results. The crucial consideration in classical probability models is that the variables considered be independent of each other.

Both the frequency and classical definitions of probability result in mathematical ratios based on counting principles. It is these ratios, the product, that comprise the subject of current controversy in the legal system. The reasons underlying the controversy are best illustrated by reference to cases in which probabilistic testimony has been in issue, and has constituted grounds for an appeal.

For example, in a California robbery case the prosecution's key witness, a mathematics professor, calculated the probability of incidence rates of characteristics which the suspects had in common with the defendants: an interracial couple driving a yellow car--a blond woman wearing her hair in a pony-tail, accompanied by a black man with a beard and mustache. Using conservative estimates, he concluded that the probability of a random couple possessing all these characteristics was 1/12,000,000. The subsequent conviction was overturned partly because the independence of the relevant characteristic events was never established, i.e., wearing a beard is not typically independent of having a mustache, although the professor treated these factors as if they were. In its opinion reversing this case on appeal, the California Supreme Court held that probability theory could not prove beyond a reasonable doubt that only one couple possessing the characteristics could be found in the area. Accordingly, characterizing the statistical evidence as "a veritable

sorcerer in our computerized society" (p. 23) that threatens to cast a spell over the trier of fact, it held that the evidence was prejudicial (People v. Collins, 1968).

A second interesting example of problems resulting from the introduction of frequency probability evidence is provided by the case of State v. Garrison, (1978) in which, following the death by strangulation of Verna Marie Martin in Tucson, Arizona, in 1976, defendant Bobby Joe Garrison was convicted of first degree murder and sentenced to life imprisonment. Key evidence proffered by the prosecution came from a specialist in forensic dentistry who testified that wounds in the deceased's breasts had points of similarity with defendant's teeth.

Despite the fact that the expert's conclusions were based on fewer than the usual ten-point comparison, he stated that the probability factor of the two sets of teeth being identical in a case similar to that was approximately eight in one million. The Arizona Supreme Court held, en banc, that the testimony was admissible and affirmed the conviction. A very instructive partial dissent addressing this issue was filed by Judge Gordon who concluded that the state's expert was "totally out of his field when the discussion turns to probability theory" (p. 568). In verifying the scientific basis for the witness's statements, the judge determined that the frequency probability the expert intended to cite was "eight in one hundred thousand," but that even this figure was misleading, since the underlying study in which the figure 8/100,000 was cited was based on the probability of distinctive tooth characteristics in Scotland, where not more than

sixty percent of the adult population over sixteen have some natural teeth. The judge went on to calculate the appropriate frequency probability, given the information used by the expert, and concluded that this was less than one in two thousand. He cautioned that to permit the expert to offer a probability figure "without a complete explanation of how the number was calculated would not only intensify the mystery surrounding pronouncements of such huge probability figures, but also foreclose the possibility of an effective defense" (p. 569).

Legal Admissibility of Probabilistic Testimony

Three threshold questions must be answered affirmatively before scientific testimony may be admitted into evidence in accordance with the federal rules of evidence: (1) whether it is relevant; (2) whether the probative value of the evidence outweighs its prejudicial value; and (3) whether this type of evidence has attained a suitable level of general acceptance in the legal and scientific communities (Frye v. United States, 1923). These standards have themselves been the source of considerable controversy, but, despite severe criticism, according to Giannelli (1983), no preferable alternatives have yet emerged from the debate. Many state courts utilize similar standards to the Frye test. For example, in New Jersey, the evidence is admissible if the proposed technique or mode of analysis has sufficient scientific basis to produce uniform and reasonably reliable results and will contribute materially to the ascertainment of the truth (State v. Cary, 1967).

The first and third prongs of the Frye test are generally easy to meet. Evidence is regarded as relevant if it tends to increase or

decrease the likelihood of the contested facts or events. Disputes over consensus in the scientific community are infrequently the basis for exclusion of scientific evidence (Giannelli, 1980; Imwinkelried, 1981; Imwinkelried, 1982). It is the second prong of inquiry, as indicated by the foregoing examples, that typically becomes the stumbling block to the admissibility of scientific evidence, i.e., "once it is determined that the expert testimony in question is competent, its probative value must be weighed against its prejudicial effect" (United States v. Amaral, 1973, p. 1152). In the case of probabilistic evidence, the phrase "prejudicial value" has come to reflect concern that the expert witness may unduly sway the jurors. For example, in excluding expert evidence about the reliability of eyewitness testimony on grounds that the prejudicial effect outweighed the probative value, one court justified its ruling as follows:

> There was a substantial risk that the credentials and the persuasive power of the expert would have had greater influence on the jury than the evidence presented at trial, thereby interfering with the jury's special role as factfinder. Scientific or expert testimony especially courts the danger of undue influence or of confusing the issues or misleading the jury because of its aura of special reliability and trustworthiness. (United States v. Collins, 1975, p. 637).

When novel scientific evidence is incriminating rather than exculpating, and the scientific foundation is challenged, courts are more likely to stress the prejudicial potential of the expert testimony and to exclude it on that basis (Monahan & Walker, 1985). For example, in New Jersey v. Cavallo, (1982), the defendant sought to offer psychological testimony on the psychological traits of a rapist. This was denied on grounds that "while juries would not always accord

excessive weight to unreliable expert testimony, there is substantial danger that they would do so, precisely because the evidence is labelled 'scientific' and 'expert'" (Monahan & Walker, 1985, p. 239).

Courts have aimed their most pointed comments at statistical evidence, but have voiced general doubt as to whether jurors can properly assess any scientific evidence (Imwinkelried, 1982). For example, the Court of Appeals in the District of Columbia asserted that jurors often attribute a "mystic infallibility" to scientific evidence (United States v. Addison, 1974). Similarly, in Reed v. State, (1978), the court held that jurors routinely overestimate the objectivity and certainty of scientific evidence. In a Washington civil case, the court held that the sophisticated scientific testimony "may exceed the ability of the lay juror to decide the facts in an informed and capable manner" (In re Boise Cascade Securities Litigation, 1978, p. 104).

Questions about juror competency to evaluate scientific expert testimony have formed the basis for motions to exclude evidence from jury trials in both civil and criminal cases, particularly when expert witnesses take opposing views. For example, in a recent criminal case a defendant charged with interstate transportation of stolen goods entered a plea of "not guilty by reason of insanity" on grounds that he was a compulsive gambler (United States v. Torniero, 1984). A motion was brought to keep all expert psychological and psychiatric testimony on this issue from the jury because the jury would be confused by the clashing opinions of the experts. Citing to Barefoot v. Estelle (1983), the Second Circuit Court of Appeals appeared to acknowledge that the issue of juror competency requires empirical validation,

holding that "psychiatric testimony should not be excluded solely as a result of an unfounded belief that 'a jury will not be able to separate the wheat from the chaff'" (United States v. Torniero, p. 734). This court also emphasized the function of the jury to evaluate conflicting evidence and reach a decision on criminal responsibility by applying society's values to the legal issues in dispute, noting:

> The Framers of the Bill of Rights expected that juries would be capable of resolving disputed issues of fact in the federal courts. Even in civil litigation, where non-perspicuous issues and abstruse evidence proliferate, we have never acknowledged a "complexity exception" to the right to a jury trial. (p. 734).

Once the threshold questions regarding the admissibility of scientific nature evidence are resolved, if the evidence **is** admitted, the controversy becomes more intriguing, as a number of judicial theories about juror competency to make rational decision have emerged, some of which are diametrically opposed. While these theories lack the characteristics of scientific formulations, they are considered opinions. Judicial opinions of juror performance in cases involving probabilistic scientific evidence are reviewed next.

Judicial Responses to Probabilistic Testimony

Judicial responses to probabilistic testimony fall into three distinct categories, with attendant but sometimes conflicting hypotheses about juror performance: (a) the evidence is not probative; (b) the evidence is prejudicial; and (c) the evidence invades the province of the jury. Based on their hypotheses about juror performance and compatible evidentiary rules, judges decide whether to

admit or exclude the evidence at trial. If the trial judge's decision is appealed, an appeals court reviews the decision in light of these hypotheses, and, after applying pre-determined standards of review or degrees of scrutiny to assess whether harmful or harmless errors have occurred, either affirms or reverses the trial court decision. The three categories of judicial responses to probabilistic testimony will be illustrated and discussed in turn.

A. Probabilistic evidence is not probative.

In some cases in which probabilistic expert testimony is not initially regarded as prejudicial simply because it is scientific, and is consequently admitted into evidence, the testimony may later be determined to be prejudicial on other grounds. First, the evidence may be regarded as prejudicial because it is not considered probative. Two facets to the "prejudicial because non-probative" line of reasoning are discernible: (a) probabilistic evidence is prejudicial because it encourages jurors to speculate about past events; and (b) probabilistic evidence is prejudicial because it is irrelevant.

1. Probabilistic evidence encourages jurors to speculate. This argument reflects concern that probabilistic evidence encourages jurors to speculate as to what happened rather than to decide the case on the basis of the particular facts before them. This viewpoint was the basis of the opinion of the court in State v. Saldana (1982), a Minnesota case in which the plaintiff sought to introduce evidence of post-rape trauma syndrome to prove that plaintiff did not fantasize the rape which was the cause of action in a civil trial. The court held:

> The jury must not decide this case on the basis of how most people react to rape or on whether Fuller's reactions were the typical reactions of a person who has been the victim of a rape. Rather, the jury must decide what happened in **this** case and whether the elements of the crime have been proved beyond a reasonable doubt. (Monahan & Walker, p. 247.)

This perspective, although it has some prominent adherents, such as Tribe (1971), has been the subject of increasing criticism from both lawyers (Imwinkelried, 1983; Tawshunsky, 1983) and from psychologists (Saks & Kidd, 1980) who argue that all factual evidence is ultimately probabilistic, since proof involves drawing inferences from the evidence (Loh, 1984).

2. Irrelevance and the "defense fallacy". A second facet of the non-probative argument, commonly known as the "defense fallacy" (Thompson, 1986), states that probabilistic evidence is not relevant because the applicable population is so large that any match between the defendant and the population cannot reasonably lead to an inference of guilt, ergo the information is prejudicial. This argument presupposes that jurors are inept at distinguishing highly probative from less probative evidence.

Many examples of how this theory may be applied in practice come from cases in which blood-type evidence is introduced. Courts are split on the admissibility of blood-type evidence in criminal cases not involving questions of paternity, with the majority of jurisdictions holding that such evidence is relevant and therefore admissible. For instance, in a leading case permitting the admissibility of blood-test results, Shanks v. State (1945), the defense moved to exclude testimony

that Type O blood found on the coat of the defendant was the same as that of the alleged victim, because 45% of the population have Type O blood, thus the evidence was too remote. The court held that:

> To exclude evidence merely because it tends to establish a possibility, rather than a probability, would produce curious results not heretofore thought of. In this case, the fact that the accused was somewhere near the scene of the crime would not, in itself, establish a probability that he was guilty, but only a possibility, yet such evidence is clearly admissible as a link in the chain. [Citations.] The admissibility of this evidence is not affected by the fact that Type O blood is common to perhaps 45% of the people of the world. It was still competent as some evidence, just as evidence of how an assailant was dressed, however conventionally, would be competent though by no means conclusive of identity. (p. 89).

In a similar case, State v. Fulton (1980), the trial court admitted testimony that bloodstains found on the defendant's shoes were type A, PGM, Hp 2-1, and that this blood type occurs in approximately 11% of the United States' population. On appeal, the defendant argued that the impact of this testimony was prejudicial, as all it proved was that thousands of people in Winston-Salem had the same blood type as the victim, thus the evidence had no relevance in the trial of this particular defendant. The court ruled that while the blood-type test was only weakly probative, it was nevertheless relevant, tending to identify the defendant as belonging to the class to which the guilty party belonged. The court further held that the fact that the weight of the evidence (its relevance), as inferred by the appellate court, was never explained to the jury, did not constitute prejudicial error.

The prevailing trend is to admit the results of blood-type tests (People v. Lindsey, 1978), but to limit their admissibility to corroboration of the defendant's whereabouts at a crucial time, unless there is additional independent evidence that (a) the accused lost

there is additional independent evidence that (a) the accused lost blood in the course of perpetrating the crime; and (b) the accused was present at the scene of the crime. When these criteria are present, the evidence is admissible to identify the defendant as the perpetrator of the crime charged. In other words, the permissible inferences to be drawn by jurors from the same evidence may vary.

From a legal standpoint, the foregoing arguments that probabilistic evidence encourages juror to speculate or that probabilistic evidence is irrelevant are both framed in terms of the lack of probative value of the probabilistic evidence. From a psychological perspective, a number of different hypotheses about juror competency were implied. The first argument presupposed that jurors would accord too much emphasis or weight to the probabilistic testimony. The theories underlying the second set of arguments are more complex. The defense assumed that jurors would accord too much weight to the "irrelevant" evidence. The courts upholding the admissibility of the evidence appeared to assume that jurors had no trouble evaluating the expert testimony and in according it appropriate weight, despite the fact that the lawyers before the court were not able to do this. Whether the subsidiary holding in <u>State v. Fulton</u> (1980) to the effect that the evidence was not prejudicial despite the absence of instructions on the weight of the evidence may also imply that jurors ignored the scientific evidence is unclear. What is clear is that this court rejected the notion that jurors accorded undue weight to the frequency probability evidence.

B. Prejudice in light of burden of proof

A second judicial theory, often espoused in excluding probabilistic evidence at trial, is that jurors will substitute the mathematical probabilities for the burden of proof and thus accord too much weight to the evidence. The Supreme Court of Minnesota noted:

> Testimony expressing opinions or conclusions in terms of statistical probabilities can make the uncertain seem all but proven, and suggest, by quantification, satisfaction of the requirement that guilt be established "beyond a reasonable doubt" (State v. Carlson, 1976, p. 176).

In this instance, the prejudice is not seen as inherent in the fact that the evidence is scientific, nor in the fact that the evidence is probabilistic, but in the interaction between the standard of proof required in criminal cases and the probability statement by the expert witness. For example, if an expert were to testify that there was a 5% probability of a mismatch, "this might be equated by some persons to a 5% probability that the defendant is innocent" (Straf, 1983, p. 229). A similar example was provided by Jonakait (1983) who cautioned that the evidentiary effect of frequency probability evidence was easily misjudged. For instance, "if an expert were to state that there was one chance in 1000 that a blood sample came from someone other than the defendant, jurors would conclude that science had established it as 99.9% certain that the defendant's blood was found at the scene" (p. 386). This inferential error has been termed "the prosecutor's fallacy".

An examination of case law reveals that it is not merely jurors who may be susceptible to this error, but that lawyers and even expert

witnesses exhibit the same vulnerability. An example of a prosecutor who committed the prosecutor's error comes from a case in which a defendant was convicted of bank robbery following a trial in which microscopic hair analysis showed defendant's hair was similar to hair found in a ski mask (United States v. Massey, 1979). The judge elicited testimony from the expert about the mathematical probability of a random match, and was told it was 1/4,500. In closing argument, the prosecutor, taking a cue from the judge, emphasized this evidence and attempted to explain its implications to the jurors:

> A handful--3 to 5 out of 2,000--that's better than 99.44 percent; it's better than Ivory Soap, if you remember the commercial. It's very, very convincing. If hair samples are microscopically identical, that is at the very least proof beyond a reasonable doubt that the unknown hair comes from the same head as the known hair. ... Just the hair sample would be proof beyond reasonable doubt because it is so convincing. (p. 681).

The appellate court found that the prosecutor, by these remarks, had confused the probability of concurrence of the identifying marks with the probability of mistaken identification of the bank robber, and ordered a retrial.

Support for the premise that jurors may confuse the probabilistic evidence with the probability of a mistaken identification comes from yet another case involving hair sample analyses and testimony that the likelihood of a random match was 1/4500. From the deliberation room, jurors sent a question to the judge asking: "Has it been established by sampling of the hair specimens that the defendant was positively proven to have been in the automobile?" (United States ex rel. DiGiacomo v. Franzen, 1982, p. 516). The trial judge, hoping to avoid reversible

error, responded that he could provide no answer to their question. The Seventh Circuit Court of Appeals upheld the conviction while acknowledging the apparent jury confusion, because the defense had not availed itself of the opportunity to challenge the probabilistic testimony via cross-examination, nor had it called its own expert. The court went on to state that it knew of no constitutional principle by means of which it could hold as improper testimony of the expert who expressed her conclusion in terms of mathematical probability:

> While the better practice may be for the court to specifically instruct the jury on the limitations of mathematical probability whenever such evidence is admitted, we have no authority to impose such a rule on the Illinois courts. (p. 19).

Whether jurors who hear probabilistic evidence in the context of a civil trial in which the burden of proof is "a preponderance of the evidence," as opposed to the context of a criminal trial in which the burden of proof is "beyond reasonable doubt," are more or less likely to substitute the qualitative probabilistic proof for the quantitative burden of proof is unknown. If we extend the logic of the court in <u>State v. Carlson</u> (1976), then jurors in civil cases should be all the more susceptible to probabilistic evidence, since the threshold for proof in civil cases is lower. However, an examination of judicial decision-making in civil cases in which statistical proofs are prevalent shows opposite results, i.e., failure to accord any weight to the probabilistic evidence may be the more likely response. For example, in civil employment discrimination cases, often tried before judges rather than juries, judges frequently ignore the statistical

testimony. Different judges may draw juxtaposing inferences from the same evidence. In disparate impact discrimination cases, in which statistical proof is the major evidence proffered, highly significant statistical proof may be regarded as conclusive on the issue of discrimination in one jurisdiction and not in another. Compare Melani v. Board of Higher Education (1983), in which the judge was persuaded by the plaintiff's statistical evidence, with Presseisen v. Swarthmore College (1978), in which the judge did not believe the statistical evidence introduced by either the plaintiff or the defendant. The statistical evidence offered in these cases typically involves regression studies somewhat more complicated than the probabilistic testimony offered in criminal cases. However, contrasting outcomes in these cases based on similar evidence serve to point out that judges are uncertain how much reliance to place upon statistical evidence (Gray, 1986) in civil cases. Judicial concern about confusion of the burden of proof and probabilistic evidence is not extended to civil cases, possibly because the perceived consequences of a verdict unsupported by the evidence in a civil case are regarded as less serious than consequences of an unsupported verdict in a criminal case.

C. Invasion of the province of the jury

Related to the issue of confusion with the burden of proof is judicial concern that probabilistic expert testimony will somehow preempt the function of the jury in deciding the ultimate factual issues. When probabilistic evidence entails the explicit use of a numerical estimate of the probability that a certain event will or will not occur, courts may exlude the evidence on grounds that jurors will

have nothing left to decide. In cases in which the crucial facts have been the subject of the testimony by the expert witness, such statements by the expert may constitute reversible error. For instance, in Jones v. State (1974), the court stated that the witness must not express on the stand an opinion of "ultimate fact", or the very fact to be decided by the jury, because to do so would be to invade the province of the jury.

Legal theorists have endorsed this viewpoint in the past. For example, Loh (1979) distinguished probabilistic evidence used to raise an inference regarding an element of a crime from probabilistic evidence used to establish an ultimate fact issue in a criminal case, such as the identity of the defendant. Use of the former he condoned, since jurors could either draw the intended inference or not. Use of probabilistic evidence to establish criminal liability, he believed should always constitute grounds for reversible error.

Congruent judicial opinions include the following: A case involving hair sample analyses was appealed when the expert stated that there was a 25,000 to one probability that the defendant was the source. The court held that this was appealable error because the expert left nothing for the jury to decide (Stogsdill v. State, 1977). Similarly, reversible error was found when an expert who testified about neutron activation analysis stated that he was 99.999% certain that the tire iron used to jimmy the door came from the defendant's car (People v. Woodward, 1964). The controversy is not confined to criminal cases. Imwinkelried has noted that in civil paternity suits, in which results of human leukocyte antigen tests are commonly used,

there is a burning dispute over the extent to which the proponent may quantify the weight of the evidence by citing a percent probability of paternity to the jury.

In summary, two sets of juxtaposing beliefs about jurors' nonuse or misuse of scientific evidence have emerged in judicial opinions: First, with regard to jurors' fact-finding function, probabilistic evidence is thought to (a) encourage jurors to speculate about what the facts may be rather than to focus on what has or has not been proven, or, conversely, (b) usurp jurors' role, leaving them no room to consider what is or is not proven. Second, when jurors do consider the probabilistic evidence in the course of determining the facts of the case, they are thought to (a) ignore the evidence (defense fallacy) or (b) confuse the evidence with the burden of proof (prosecutor's fallacy). All four perspectives are concerned with the weight jurors accord to the evidence.

The Dominant Issue is the Weight of Probabilistic Evidence

The threshold issue of admissibility *per se* is rarely the focus of attention in cases involving probabilistic evidence. Those who favor admitting probabilistic scientific evidence argue that since the jury is free to disregard the evidence entirely, or to use it constructively, the jury is still in control (Abney, 1986). Others who favor the liberal admission of scientific evidence point out that the procedural safeguards built into the legal system are adequate to prevent jurors from misusing scientific evidence. Precisely what lawyers and experts should strive to achieve to ensure that the evidence is accorded appropriate weight is unknown. Central questions

to be answered in this regard were outlined by Graham (1983) at a national conference of lawyers and scientists:

> Lack of mutual understanding between scientists and lawyers accounts for inability of lawyers and scientists to explain to jurors the true import of the evidence. Perhaps this is why we really do not know if jurors place undue weight on evidence that happens to be called scientific, although the prevailing assumption is that they do. Again, we do not know the relative effectiveness of experts who carefully explain the evidence and sincerely try to educate the jury, and experts who simplify and in the process, perhaps distort the evidence in an effort to communicate easily with the jury. Which are more persuasive is an open question. (p. 232)

According to Imwinkelreid (1981), a new phase in the evolution of scientific evidence has commenced--a phase in which questions about the weight of scientific evidence will predominate, rather than its admissibility. Already, some courts have foreseen this trend, stating that "once evidence rises to the status of scientific principle, the question shifts to the weight to be accorded to the testimony rather than its admissibility" (United States v. Hulen, 1977, p. 278).

Existing Constraints to Avoid Prejudice

In a discussion of four cases in which probabilistic evidence was introduced in cases tried before a judge, Bar-Hillel (1984) observed that the faulty probability calculations on the part of the judges may have brought about faulty judgments, but she elected to reserve judgment on the role which quantitative probabilistic analyses do and should play in courts. Similarly, judgment about the competency of jurors should be reserved until there is better evidence available regarding their performance. It is possible that the existing constraints in the judicial system are adequate to prevent or

effectively minimize juror errors. Perhaps all that is necessary to minimize error is more effective and emphatic use of these constraints, such as cross-examination of expert witnesses, use of an opposing expert, explanation by lawyers during closing arguments, instructions from the judge to the jury, and illustrative exhibits.

Effective cross-examination of the expert is a safeguard of the advocacy system of justice to ensure that jurors are exposed to issues underlying the opinion and to alternate viewpoints. Cross-examination of experts can instruct jurors on the proper inferences to draw from the probabilistic testimony. For example, in an extensive critique of the response of courts to hair sample evidence, Tawshunsky (1983) recommended that during cross-examination of the expert, attorneys should stress the proper significance of frequency probabilities, i.e., that hundreds of people in a metropolitan area might have the hair samples to match those of a defendant, and that the hair samples cannot prove that the defendant committed the crime.

Tawshunsky (1983) also advised attorneys to explain the appropriate inferences of the evidence in the closing arguments. For example, they should clarify that the hair sample analyses or other frequency probabilities lend support to the proposition that the defendant was the perpetrator, and that a smaller probability of a random match lends more weight to the evidence. Their goal should be to put the study underlying the probability statement into proper perspective.

Others have recommended that judges expand their role in ensuring the jurors accord the proper weight to scientific evidence. Moenssens (1983) suggested that particularly when novel scientific evidence is

admitted, the judge should provide a special set of instructions that go beyond the usual cautions on uses of expert testimony. Special instructions should outline the fact that there are opposing views about the validity of the scientific evidence. Jonakait (1983) believed an instruction on the limited probative value would be appropriate when frequency probability evidence is introduced, but noted that there are no cases in which use of such an instruction has been reported. After an exhaustive review of current approaches to the uses of frequency probabilities in criminal cases, he concluded that "no workable method exists for effective and fair introduction of blood marker probability statistics at trial" (p. 421).

Legal analyses of the issue such as that by Jonakait (1983) are generally based on reviews of legal precedent, and do not anticipate an empirical solution. However, the question of juror competency raises several sociological and psychological issues which can be addressed empirically. For example, there is no empirical evidence that judges perform any better than juries in dealing with the issues raised by the inclusion of probabilistic scientific evidence. Similarly, recommendations that existing constraints in the advocacy system are adequate if they are put to better use provide many opportunities for future research.

Lempert (1981) examined the judicial crisis on whether to permit jurors to try cases involving complex evidence, and in so doing, isolated the issues that bear further investigation, one of which was whether jurors consider expert testimony in reaching a verdict. He enumerated four techniques for evaluation of the problem and empirical

assessment before this issue is decided:

(1) study of archival records on cases in which complex evidence has been presented;

(2) interviews of judges and jurors;

(3) use of shadow juries in actual trials;

(4) jury simulation studies.

Lempert suggested that simulation studies are best suited to assess what jurors do well, what they do poorly, and to test methods of presenting evidence to jurors and jury comprehension.

In a more recent review of the use of jurors in cases in which complex evidence is presented, Austin (1984) also advocated empirical evaluation of the questions posed by the introduction of complex evidence in jury trials. He noted that if jurors fail to make rational decisions in cases involving complex evidence, (a) the objectives of the law may go unfulfilled; (b) verdicts will be unpredictable; and (c) procedural rules for fair trial will be useless. Of course, these same concerns are equally applicable to situations in which judges fail to understand/rationally evaluate complex evidence.

Austin (1984) went on to summarize reasons offered in support of a preference for either a judge or a jury trial when complex evidence is involved. Scrutiny of these reasons reveals a host of unaddressed empirical questions. For example, adherents of the movement to abandon the jury system emphasize the viewpoint that, by comparison, a judge:

a. performs better at separating relevant from superfluous issues;

b. can better translate and distill expert testimony;

c. can correctly apply legal standards.

Among the advantages of a judicial decision a number of more flexible procedures available to the judge were also cited. For example, a judge:

d. can review transcripts daily;

e. is under no pressure to render an immediate verdict;

f. can reopen trial for clarification or for new evidence;

g. can appoint experts to explain the evidence.

An examination of past practices provides scant evidence to suggest that judges avail themselves of the opportunity to implement these procedures (Panel on Statistical Assessments as Evidence in the Courts, 1985), available in both civil and criminal cases since 1975.

Proponents of the jury system, by contrast, argue that jury trials are preferable when complex evidence is involved because:

a. collective wisdom is synergetic, more than the sum of the individual parts;

b. collective experience ensures better evaluation of the credibility of witnesses;

c. many jurors have technical expertise;

d. litigants are insulated from biased judges;

e. this system forces lawyers to be efficient and better communicators.

Empirical resolution of the issues posed by the increasing introduction of probabilistic evidence in litigation will have implications not only for decisions to be made regarding a particular

case which involves this sort of evidence, but also for decisions of Constitutional magnitude. By the Sixth Amendment to the U.S. Constitution, defendants in criminal trials are guaranteed a right to a public trial by an impartial jury. Courts are unlikely to grant a complexity exception to the right to a jury trial in criminal cases (United States v. Torniero, 1984). Insofar as civil cases are concerned, the movement to implement a complexity exception to the Seventh Amendment right to a jury trial has gathered some momentum, although, according to Arnold (1980), historical support for this exception is not well-founded. To the extent that jurors' judgments in cases involving probabilistic evidence may be biased, the Constitutional guarantee to an impartial decision may remain unfulfilled.

Conclusion

The foregoing analysis of case law on the ability of jurors to understand and apply probabilistic evidence indicates that judges engage in a great deal of speculation about jury behavior and are fairly skeptical about jurors' abilities in this regard. Judges have theorized that jurors are unable to make appropriate decisions because of the following specific biases: (1) overbelief in science; (2) overbelief in expert witnesses; (3) inability to draw the appropriate inference from and to accord proper weight to the evidence; (4) confusion of probabilities with the burden of proof; (5) failure to understand scientific evidence. Judicial opinions of this nature have fueled the debate over the viability of the jury system and intensifed

the need for empirical resolution to the issue. Notwithstanding the considered judicial opinions about juror performance, it is nonetheless difficult to predict precisely what evidence is too complex for jurors to render a rational verdict. For more substantive answers to this question, psychological findings that bear on these issues are reviewed in the next chapter.

CHAPTER TWO

PSYCHOLOGICAL PERSPECTIVES ON MOCK-JURORS' INFERENCES

Introduction

The purpose of this chapter is to examine jury decision-making involving probabilistic expert testimony in the light of insights from psychological research. Pychological research on cognitive processes of individuals who make choices in conditions of uncertainty, as well as research on knowledge structures in ambiguous situations, and research on attitudes and decision-making will be examined with a view to showing how those areas of inquiry bear on the biases which judges theorize jurors exhibit in cases involving probabilistic scientific evidence. Some individual biases, known as traits, are long-term predispositions. This chapter will also explore some techniques which may lessen the impact of juror biases. Other less enduring biases or moods, temporarily induced, are state biases. In recent years, psychologists have searched for ways to minimize the effects on judgment of state and trait biases. Kaplan and Schersching (1980) have noted that any technique which increases the amount of information effectively taken into account in judgment formation would arguably lessen bias. Of course, if Jensen's (1985) position on bias in mental testing were applied, and mock-jurors' initial selection of information to consider was deemed biased, these techniques would have less impact.

Various theoretical models or frameworks have been proposed in the course of research on contingent decision behavior, including the cost/benefit approach, the perceptual model, and rule-based production systems, reviewed and evaluated by Payne (1982). Within each model,

some effort is made to distinguish between task effects and context effects, though sometimes the terms have overlapped. With respect to the task of juror decision-making, a number of models of processes and their components have also been posited. For example, based on comparisons in the way in which two different juries responded to the same complex trial, Austin (1984) described a model in which the influence of personal beliefs, value judgments and other biases operated to produce coping mechanisms to deal with task variables such as information conflicting with pre-existing juror biases. He also noted that some coping mechanisms or strategies were developed in response to procedural or contextual variables. Among the coping mechanisms he listed were jurors' tendencies to ignore, distort or minimize the value of the complex or technical evidence.

In other models of juror decision-making, task and context variables have been more rigorously distinguished. For example, one useful frame of reference for examining juror decisions is a tripartite information-integration model proposed by Kaplan and Schersching (1980) in examining the effects of juror biases on juror judgments. The three components are (a) task variables, described as "information about the judged object"; (b) context variables, described as "situational demands"; and (c) moderator or individual difference variables, described as "the personality of the decision-maker." Note that in the Kaplan and Schersching model, stable personal dispositions are labeled as trait biases, while transient dispositions are labeled as state biases. State biases are usually situationally induced, creating a margin of overlap between the second and third components of the

model:

> 1. Information about the judged object. In a trial this could consist of evidential testimony, which possesses a scale value for guiltiness, as well as weight. Factors such as witness credibility, logical consistency, and so on, would affect the latter.
>
> 2. Situational demands. Here would be included deliberation effects, legal restrictions, time demands, needs of society, and peer (other juror) pressures.
>
> 3. Personality of the decision maker. The juror's pretrial biases, whether specific to the defendant or general toward all accused persons, and whether relatively permanent (trait characteristics) or transient (state) characteristics) enter here. (Kaplan & Schersching, 1980, p. 151).

In reviewing the psychological literature which bears on the question of juror competency in cases involving probabilistic scientific evidence, this model will be used as a frame of reference to examine variables which may operate to influence each of its three components. However the major focus of this discussion is on the first component, i.e., the influence of variations in the nature of the probabilistic information presented to jurors.

Cognitive Assessments of Probabilistic Evidence

A. Decision-making and non-statistical heuristics

A well-developed body of research has established that humans engage in a number of judgmental heuristics to reduce the cognitive strain of complex information-processing tasks (Tversky & Kahneman, 1974; Nisbett & Ross, 1980; Christensen-Szalanski, 1980). Use of judgmental heuristics may systematically bias the perceptions, structuring, and processing of task-relevant information (Payne, 1982).

For example, humans selectively discount or ignore information that is difficult to deal with by favoring a readily available interpretation of a situation.

The task which confronts jurors can readily be analogized to the experimental procedures used in studies of judgmental heuristics. In some recent laboratory studies by Tversky and Kahneman (1983) scenarios more typical of the experimental materials used in jury simulation studies were employed, strengthening the applicability of this body of research to juror decision-making. For example, as part of a series of studies on problems (varying in transparency) which produce the representative conjunction fallacy, subjects were given a brief written description of an individual who was convicted for one crime, and were then asked to rank order a series of possible results of a second ongoing criminal investigation of the same individual. Half the subjects received as one of the possible outcomes: "Mr. P. killed one of his employees." The other half of the subjects were told: "Mr. P. killed one of his employees to prevent him from talking to the police." The experimenters hypothesized that inclusion of a plausible but nonobvious motive in the second version would increase the perceived likelihood of the event, despite the violation of the conjunction rule. The data confirmed their expectation. However, it is noteworthy that the second statement is an elaborated proposition. In view of the finding that elaboration enhances the memorability of items, the reported effect may be attributable to the enhanced salience of this item as a consequence of facilitation through this memorial strategy.

In related research, Tversky and Kahneman (1983) found support for

the tendency of subjects to rate conjunctive events more likely than the occurrence of sub-parts when the additional information appears representative of the actor's disposition. They reported three circumstances in which mentioning a cause or motive increases the perceived likelihood of an event if the motive (a) offers a reasonable explanation of the target event; (b) appears fairly likely on its own; (c) is nonobvious in the sense that it does not immediately come to mind when the target event is mentioned. Once again, these three circumstances enumerated are characteristic examples of elaborations encouraged to enhance memory of propositions, thus enhanced memorability cannot be ruled out as a possible explanation for the findings.

Noting that vulnerability to conjunction errors is a robust phenomenon when using scenarios to assess probability judgments, Tversky and Kahneman (1983) drew a parallel between these findings and what takes place in a trial. For example, "an attorney who fills in guesses about unknown facts, such as motive or mode of operation, may strengthen a case by improving its coherence, although such additions can only lower probability" (p. 308). The researchers emphasized the need to study decision-making based on limited information, such as occurs when certain facts are selected for presentation in the course of a trial:

> The implications of the psychology of judgment to the evaluation of evidence deserve careful study because the outcomes of many trials depend on the ability of a judge or jury to make intuitive judgments on the basis of partial and fallible data (Tversky & Kahneman, 1983, p. 307).

Research of a variety of information processing and integration tasks has produced several theories about the repeated finding that people tend to ignore some salient information when making a decision. Christen-Szalanski (1980) and others posited the cognitive strategies which people use vary with the task at hand, and are a function of the costs and benefits of each. Thus, the benefit of the most accurate strategy may be ignored if the mental costs and effort of processing are high.

Social psychologists studying attributions and inferences which people make about the conduct of others have found that people have a strong predisposition to attribute causes, either internal or external, to the behavior of others in an effort to explain it (Zadny & Gerard, 1974; Heider, 1958). More recently, psychologists have begun to apply these theories to the legal domain (Devine & Ostrom, 1985; Hastie, 1983; Tversky & Kahneman, 1980; Coates & Penrod, 1980). One implication of the attribution literature for jury decisions is that in cases in which there are plausible human motives or environmental factors to explain the facts, jurors are likely to impute more weight to these factors than they will to statistical proofs. Jurors may be predisposed to ignore or undervalue the importance of probabilistic scientific evidence because this type of information does not comport well with their preference for making decisions based on inferred causal hypotheses about a defendant's conduct.

One of the issues that has dominated the social cognition literature is when the discounting of information takes place, i.e., does this occur at the time the information is presented, or at some

later stage is the information recalled, intact, for evaluation and then discounted at the time a verdict is rendered? Devine and Ostrom (1985) challenged the assumption that jurors suspend evaluation of testimony until the end of the trial. Their research shows that jurors actively evaluate testimony as it is received and continue to modify the constructed story or script throughout the trial. Even when subjects were told that their task was to recall certain items, not to make judgments or reach a verdict, they nevertheless demonstrated the continuing evaluation strategy during story construction. These researchers believe that this strategy is spontaneously employed whenever stimulus materials such as trial materials are presented. They concluded that any discounting of the evidence, for instance, because of source credibility, etc., occurs at the integration stage, as jurors generate an integrated cognitive representation of the story or trial events. Evaluative measures revealed that mock-jurors imparted very different meaning to and formed different cognitive representations of the same testimonial items, leading them to different verdicts. Devine and Ostrom concluded that evaluation of the credibility of even one witness could exert a substantial impact on jurors' verdicts.

A third line of research, which explored other factors which influence the way in which the people remember events, examined the impact of vivid versus pallid information on decisions. Results of this research led psychologists to hypothesize that jurors may selectively ignore or discount information that is probabilistic because it is so dull in comparison with anecdotal testimony (Saks &

Kidd, 1980). Thus, unlike judges who fear that probabilistic information will eclipse all other testimony (Loh, 1979), psychologists have hypothesized that jurors are more likely to ignore probabilistic information than to accord it undue weight. This hypothesis is consistent with the finding that judges ignore probabilistic evidence (Presseisen v. Swarthmore College Board of Higher Education, 1983). There are some, albeit a small number of studies which provide support for this hypothesis. This research is reviewed below.

B. Do jurors ignore scientific probabilistic testimony?

While most experimental studies of jury decision-making have failed to detect systematic predictors of jury behavior, results indicate that the influence of the evidence on decision-making is strong (Hastie, Penrod & Pennington, 1983). This general conclusion, somewhat encouraging in the face of current judicial pessimism regarding juror competency in complex litigation, does not differentiate between scientific, circumstantial evidence offered though an expert witness and eyewitness testimony. Do jurors draw a distinction between these classes of qualitative evidence and attribute more weight to one than to the other?

An early field study conducted as part of the University of Chicago Jury Project (Broeder, 1959) shed some light on this matter. Included in post-trial interviews of approximately 1500 jurors who had served on 213 different criminal trials was the question: "Would you convict on circumstantial evidence?" The researchers reported that very few jurors understood what circumstantial evidence was. More recently, other researchers included questions along these lines in an

empirical study, and reported that unless jurors are explicitly instructed on the meaning of circumstantial evidence, they do not apply this concept, and tend to draw the opposite inferences about its use from that intended by law. Specifically, Buchanan, Pryor, Taylor and Strawn (1978) found that a common misperception by jurors is that circumstantial evidence does not constitute legal evidence. Consequently, jurors believe circumstantial evidence should not be taken into consideration in reaching a verdict and they often discount or ignore the value of circumstantial evidence.

Confirming evidence for this finding comes from some archival data gathered from courthouse records showing that even explicit judicial instructions to consider circumstantial evidence are confusing to jurors who are uncertain what weight to accord to circumstantial as opposed to direct evidence. Severance and Loftus (1982) reported that in a case involving circumstantial evidence, jurors sent a question from the deliberation room asking whether an instruction that they "may draw an inference" from this evidence meant that they "must" draw such an inference. The judge's response was that the answer was "unknown".

Information about the way in which jurors tend to view scientific circumstantial evidence comes from the report of a series of post-trial interviews conducted by Saks and Van Duizend (1983). They questioned judges, lawyers, expert witnesses and jurors who participated in nine cases in which scientific evidence was crucial. The researchers reported a number of general conclusions that bear on this issue. Included among these were participants' impressions that statistics were less persuasive than eyewitnesses; that anecdotal data

were more valuable than expert data; but that scientific evidence could be very credible. The researchers noted that while jurors tended to be skeptical of all expert witnesses, they also had a tendency to be impressed by "big words", and often failed to understand the substance of the experts' scientific testimony. Saks and Van Duizend (1983) concluded that whether jurors believe expert testimony **at all** was an important, unresolved empirical question.

The handful of empirical studies that have addressed this issue indicate that jurors are not as likely to overestimate the value of scientific evidence as judges imagine. For example, Loftus (1980) conducted an experiment to determine whether mock jurors who read about a bad check case were more willing to convict a defendant identified on the basis of lay testimony or on the basis of fingerprint analysis, handwriting comparisons or a polygraph test. The mock jurors were most likely to convict on the basis of the lay testimony (78%) as opposed to fingerprint comparisons (70%), lie-detection (53%) or expert analysis of handwriting samples (34%).

Similar research by Markwart and Lynch (1979) on the relative weight attributed to polygraph tests revealed that only 14.5% of the mock-jurors believed the lie-detector evidence was more significant than lay testimony. In another simulation study in which polygraph evidence was presented, jurors frequently returned verdicts inconsistent with the polygraph evidence (Cavoukian & Heslegrave, 1980). Findings that jurors are less likely to give scientific evidence undue weight than most courts presume are consistent with the conclusion of Taylor and Thompson (1982) in reviewing research on

information processing and the vividness effect. While noting that in general there is little evidence in support of the vividness effect, they theorized that subjects might underuse base-rate information in statistical presentations. This conclusion was based in part on findings such as those by Chaiken and Eagly (1976) to the effect that case histories carry more weight than statistics. Accordingly, a hypothesis set forth by Jaffee (1979) bears some examination. He predicted that where circumstantial evidence is probabilistic, and direct eyewitness testimony is available, jurors will place more emphasis on the latter, because perceptions of the past are preferable to estimates of what took place in the past. What jurors may do in reaching a decision involving weak direct evidence plus strong (highly probative) circumstantial evidence is less easy to predict.

The focus of inquiry in the foregoing studies was jurors' preferences for eyewitness testimony when contrasted with complementary or contradictory scientific evidence. Thus, these studies do not adequately address the question posed by Saks and Van Duizend (1983) concerning jurors' tendency to ignore probabilistic scientific evidence. One exploratory study designed specifically to assess how jurors use probabilistic scientific evidence was conducted by Thompson (1984), using a written summary of a trial involving a bank robbery by a red-haired male, wearing a ski-mask, based on United States v. Massey, (1979). In this experiment, subjects provided a preliminary estimate of the defendant's guilt, and were then presented with probabilistic information about forensic analyses of matching hair samples taken from a ski-mask and from the defendant's head. After

reviewing this additional information, subjects provided a second estimate of the defendant's guilt. Despite the obvious salience of the probabilistic information, as many as 12% of the subjects failed to revise their first estimates in light of the new incriminating information. Subjects who underuse or ignore probative probabilistic information exhibit what Thompson calls "the defense fallacy". These subjects "assume that a match between the defendant and perpetrator with respect to some characteristic is irrelevant because at best, it shows that the defendant and the perpetrator are members of the same large group" (Thompson, 1984, p. 2). Because of the experimental procedures compelling subjects to commit themselves to a position before considering the probabilistic testimony, the possibility of cognitive dissonance cannot be excluded as an explanation for this result. Of the remaining subjects whose probability estimates were revised following presentation of the additional incriminating information, 13% exhibited the opposite tendency, to overweight the information, which Thompson calls "the prosecutor's fallacy". These subjects deducted the incidence rate of 2% from 100 to provide a 98% estimate of guilt on the part of the defendant, regardless of the prior probability of guilt. These same lines of reasoning are reflected in several legal appeals brought on grounds that evidence is irrelevant (defense error), or on grounds that lawyers or experts have committed reversible error by overstating the implications of probabilistic testimony (prosecutor's error).

Evidence of variations in judges' responses to probabilistic scientific evidence is provided in the form of some archival data

concerning 44 trademark infringement cases which came to trial between 1957 and 1983, in which expert testimony regarding 67 different surveys was proffered. These data showed that the weight accorded to the expert testimony by the judges hearing the cases varied considerably (Jacoby, 1985). Evidence concerning 7 surveys was declared inadmissible. When the survey evidence was admitted, for 65% of the surveys, judges held the evidence to be irrelevant, to have no bearing on the case, or to have little if any weight. Thus, there is some indication that judges, too, are susceptible to the defense fallacy. Moderate consideration, or "some weight" was accorded to 10% of the surveys. Only in 20% of the cases did the survey evidence receive substantial weight, reflected in judicial opinions describing the evidence as "compelling", "particularly persuasive", or of "considerable weight". Susceptibility of judges to the prosecutor's fallacy is more difficult to assess from these data since these were civil trials requiring proof by a preponderance of the evidence, and there are no measures of how much weight was accorded to evidence characterized as particularly persuasive or compelling. While none of these cases involved jury trials they provided important information about the general tendency on the part of the judiciary to ignore or undervalue the weight of scientific, circumstantial evidence.

Results of the foregoing studies provide some support for both the psychologists' hypothesis that jurors will ignore probabilistic testimony, and the judges' hypothesis that jurors will attribute too much weight to probabilistic evidence. However, the results also show that mock-jurors' responses are somewhat more varied than either

psychologists or judges have generally predicted. Three broad classes of responses emerged: (a) underuse of the information; (b) overuse of the information; and (c) use the information to some degree between those two extremes. Because of the exaggerated salience of the probabilistic information in the one empirical study by Thompson (1984), some caution about the generalizability of the results is advisable. A set of facts in which the scientific evidence is contested, or in which it is juxtaposed with non-scientific evidence may elicit different evaluative processes and different results.

C. <u>**Do jurors assessing probabilistic evidence make systematic errors**</u>?

The Thompson (1984) study revealed that a certain percentage of jurors were susceptible to one of two judgmental errors or biases in evaluating probabilistic evidence pertinent to one criminal case. Whether these biases generalize to other fact patterns is unknown. Implicit in judge's theories of juror performance is the notion that juror biases generalize to a variety of factual circumstances. For example, jurors are commonly believed to exhibit a prosecutorial bias in making decisions involving probability information which should simply permit the inference that a short-list of suspects can be constructed. The smaller the list, according to George (1981), the greater should be the inferred weight of the probabilistic testimony. Jonakait (1983) expressed doubt that jurors understand that frequency evidence merely defines a class of potential suspects. Cognitive scientists and jury researchers alike have begun to investigate whether jurors are susceptible to the prosecutor's fallacy and other errors when making decisions involving probabilistic facts. One aspect of

this research is an examination of conditions that elicit or impede these biases. In analyzing judgmental errors, psychologists have acknowledged that the dividing line between misunderstandings and fallacious reasoning is not always clear, but that it is useful and necessary to distinguish communication failures which produce errors that are verbal or technical, from judgmental fallacies, which are non-trivial, conceptual errors (Tversky & Kanheman, 1983).

Some information about mock-jurors' abilities to evaluate and integrate probabilistic information with other more qualitative evidence presented in a jury simulation study concerned a robbery trial of a defendant who cut himself on a broken window (Faigman, 1983). Controverted facts were presented by five witnesses, including a physician who testified that the suspect's blood sample matched a sample taken from the scene of the crime. Independent variables included the frequency of the matching blood-type in the population (40%, 20% or 5%). Subjects assigned to the 5% condition gave significantly more weight to the blood group evidence than did subjects in the other two groups, but even they underutilized the evidence in comparison with a Bayesian model. With the exception of subjects in the 5% group who stated a probability of guilt prior to the presentation of the statistical information, subjects "virtually ignored the probabilistic testimony." In the 40% and 20% frequency conditions, subjects failed to readjust their estimates to take the probabilistic information into consideration in accordance with Bayesian norms. Faigman concluded that subjects may give some weight to extreme figures, such as 5%, and little or no weight to modest

figures, but that they do not discriminate between the two in any refined manner. The finding that even where subjects used the probabilistic information, they underutilized it in comparison with the Bayesian model refutes the hypothesis that jurors are mesmerized by mathematical evidence. Rather, this study lends support to the theory that jurors are reluctant to use statistical information when making causal attributions.

Intrigued by the finding that jurors exhibit both prosecutorial and defense-oriented errors in processing probabilistic information, Thompson (1985) proceeded to study the persuasive appeal of these fallacious statistical arguments to mock-jurors. Subjects read a brief pre-trial case summary in which weak circumstantial evidence was presented, from which a police detective concluded there was a .10 likelikood that the suspect was guilty. Then, subjects provided an estimate of defendant's guilt. Additional information probative of guilt was presented to the subjects regarding forensic tests indicating that the likelihood of obtaining matching evidence by chance alone was 1%, i.e., only 1% of the relevant population possessed the suspect characteristic possessed by the defendant. After receiving this additional information, subjects read two brief arguments on the meaning of the evidence. Half the subjects first read a summary of the prosecutor's fallacy in which the defendant was described as 99% guilty, and then read a defense fallacy argument to the effect that the frequency probabilities were irrelevant because so many individuals had the same suspect characteristics. The remaining subjects received the materials in reverse order.

Thompson (1984) reported that subjects were susceptible to the fallacious arguments, did not notice the fallacious reasoning, and adopted the arguments themselves. The persuasive appeal of the arguments depended upon which argument was presented first. Subjects who received the argument favoring the prosecution first were more likely to adopt this line of reasoning and render a verdict in which the likelihood of the defendant's guilt was rated as .99. Subjects who first read the defense argument were most susceptible to its appeal. This line of reasoning was adopted by 68.5% of these subjects, who did not rate the defendant as any more likely to be guilty than they did before receiving the probabilistic evidence.

One weakness of the study is that cognitive dissonance may partially explain the results. Because the experimental procedures required subjects to estimate the likelihood of the defendant's guilt after the presentation of the first argument, subjects may have been reluctant to modify their decisions after reading the countervailing argument for fear that they would appear too indecisive, or too readily dissuaded from their previous opinions. Their commitment to a decision after the first argument may also have set up a memory process rendering recall of facts in support of prior decision more available. Prior research has shown that people have better recall for arguments in support of their decisions than for propositions which contradict their decision. A third psychological phenomenon may bear on these results is hindsight bias. In other words, these results do not permit a clear assessment of the impact of the probabilistic evidence per se. Accordingly, errors may be attributable to judgmental fallacies,

experimental procedures, or communication failure. Further research is needed to clarify this issue.

This study does indicate that subjects may easily be misled, when probabilistic information is in issue. The finding that college students are vulnerable to fallacious pseudo-statistical arguments may be interpreted by the judiciary as confirmation that jurors have problems evaluating probabilistic evidence. Assuming the propensity to be misled by fallacious statistical arguments proves to be a robust phenomenon, the corollary issue of factors instrumental in minimizing this susceptibility will assume more prominence. Perhaps jurors can be inoculated against this pitfall just as inoculation has proved effective in debiasing subjects in persuasive communication studies, according to Pryor, Buchanan and Strawn (1980).

D. Do jurors overbelieve witnesses who present probabilistic evidence?

As a practical matter, in court, jurors' responses to the substance of the scientific evidence cannot readily be distinguished from their responses to the source of that information, the expert witness. From an experimental standpoint, this distinction may be instrumental in evaluating the merit of the conflicting hypotheses about juror competency. Despite fears expressed by judges that experts are vested with a special badge of credibility causing jurors to be reluctant about questioning their testimony, evidence gathered on jurors' responses to expert witnesses from a number of sources indicates that there may be little support for this viewpoint. Austin (1982) interviewed two groups of jurors who served on a particularly complex antitrust trial that lasted three months, as part of a study on juror

competency in complex litigation. He reported that the jurors were skeptical about the experts who testified on electronics and economics. He attributed their negative response to the experts to a natural distrust of the unfamiliar. Of course, survey data cannot exclude the possibility that jurors' opinions were clouded by extra-legal factors such as the speech patterns and attractiveness of the witness or the jurors' post-hoc rationalizations, selective retrieval of information from memory, etc., all of which have previously been shown to influence jurors' decisions.

The question of how much weight is accorded to an expert witness was examined by Faigman (1983) in the study involving a defendant who cut himself during the commission of a robbery. Subjects' ratings of the credibility of the expert witness, a physician, who presented the critical probabilistic testimony, were compared with ratings given to four other witnesses, as follows: (a) an unreliable eyewitness (b) the investigating police-officer; (c) a second expert, a statistician (who did not present probabilistic information); and (d) the defendant. Significantly higher ratings were reported for the testimony of the physician than for all the other witnesses. The police officer received significantly higher ratings than did the statistician, the second expert. The statistician received the same rating as the eyewitness who admitted he had been drinking the night of the burglary.

Findings such as these indicate that jurors accorded more weight to the expert who presented probabilistic information pertinent to the factual issue to be decided than to the expert who presented no new information but who explained how to use the probabilistic information.

Previously, Saks and Kidd (1980) have argued that:

> An expert who reports only scientific or statistical data will have less impact than an expert who reports a case study, relates a compelling experience or offers anecdotal evidence, since the latter is more concrete, vivid, and emotion-arousing, thus will probably be more available to jurors during deliberation. (p. 137.)

While the Faigman (1983) study did not require subjects to evaluate anecdotal versus statistical information presented by expert witnesses, it does illustrate that an expert witness, in this case a physician, who reports only scientific or statistical data, will not be ignored if the information is probative. However, these findings are not informative regarding the weight accorded to expert testimony in reaching a verdict, because an expert may be highly credible, yet his or her testimony may be perceived as irrelevant to the factual issue to be determined by the factfinder. Faigman found no significant changes in the conviction rate before versus after the physician testified. But jurors' subjective estimates of the likelihood that the blood found at the scene of the crime was the defendant's increased following the presentation of the information about the physician. One possible explanation for this apparent incongruity in results is that jurors did not attribute enough weight to the probabilistic testimony to find the defendant guilty "beyond a reasonable doubt". In other words, the jurors may have taken the information into consideration, as indicated by the change in subjective probability ratings, but their interpretation of the burden of proof may have established the decision threshold at a level higher than that attained by the proof of guilt. If this is true, then verdict alone is a dependent measure insensitive

to changes in the weight of the probabilistic evidence, and Faigman's conclusion that mock-jurors ignored the probabilistic evidence may be unwarranted.

None of the empirical studies to date have explored jurors' responses to a "battle of the experts". For recent reviews of jurors' responses to expert testimony, see Greene, Schooler and Loftus (1985), and Law and Human Behavior (1986). The dominant theory is that jury confusion will increase when one expert's testimony is controverted by an opposing expert. This may cause some jurors to discount the testimony of both experts and base their decisions on irrelevant factors or pre-existing biases, rather on than the available evidence before them. Studies of mock-juror decision-making in situations of varying levels of uncertainty may shed some light on this issue, particularly studies in which the interaction of subject biases with uncertain information in a decision-making task has been the focus.

Kaplan and Scherschching (1980) conducted such an experiment, and found, not surprisingly, that mock-jurors respond differently when facts of mixed evidentiary value are presented, i.e., when evidence is controverted, than when it is uncontroverted. Half the subject-jurors were presented with a factual scenario in which the evidence was balanced between the parties (mixed), and the remainder of the subjects received uncontested evidence. Mock-jurors pre-tested on a harsh-lenient attitude dimension read a trial summary and rendered a verdict, listing evidence which influenced their verdicts. Their verbal answers were rated by two independent scorers as either incriminating or exonerating in value. When the evidence was contested, lenient

subjects cited less incriminating and more exonerating facts. When the evidence was uncontested, harsh and lenient subjects cited approximately the same proportion of exonerating facts. Uncontested facts received more weight from all jurors. When facts were uncontested, subject biases played a less important role in the decisions. When facts had less weight (i.e., were contested), the role of subject biases increased. The researchers observed that traditional trial procedures seem designed to enhance biasing effects. They concluded by asking whether a given manner of presenting evidence could lessen bias effects when facts are controverted. This question is applicable to the presentation of any evidence that may interact with juror biases.

E. Are jurors sensitive to the variations in probabilistic evidence?

One of the most robust findings in empirical research on information-processing and decision-making is that judgment and choice are sensitive to changes in the demands of the task, even apparently minor changes (Payne, 1982; Einhorn & Hogarth, 1981). Reversals of preferences when subtle modifications of the response mode are made have been well-documented. For example, Lichtenstein & Slovic, (1971) found that subjects' responses changed when they were asked for bids as opposed to choosing between two alternatives. Coombs, Donnell and Kirk (1978) noted that instructions to select or to reject tend to focus attention or to change the salience of the various components of an option, or the order in which they are processed.

Lichtenstein, Fischhoff and Phillips (1982) conducted a review of calibration literature, and reported that subjects are typically

overconfident when making predictions about general knowledge items of moderate or extreme difficulty. Overconfidence was most extreme with tasks of great difficulty. Even experts, such as weather forecasters, overestimated the probability of rain, with or without computerized feedback. When probabilistic information is involved, subjects have also been shown to to be insensitive to variations in the reliability of probabilistic information (Kahneman & Tversky, 1973) or to underestimate their impact. This issue was further investigated by Kruglanski, Friedland and Farkash (1984), who reported that subjects did use reliability information properly when its applicability was apparent. They argued as a result of this contrary finding that talking about people's statistical intuitions in general is unwarranted, and that since specific conditions may determine the degree to which statistical notions are perceived as situationally applicable, most errors are errors of application rather than errors of comprehension. This perspective is encouraging for legal practitioners and judicial administrators, first because they have the opportunity to make the applicability of the information more pointed, and second because appropriate remedies can more readily be fashioned.

Whether a trial involving probabilistic scientific testimony is perceived by jurors as an applicable situation in which to consider reliability information is unknown. The reliability of scientific evidence presented in court varies considerably, depending upon the procedures used, variations in frequencies pertinent to different populations, applicable confidence intervals, etc. Sperlich has (1985) noted that jurors need assistance in data, generalizations and opinions

which meet scientific standards from those that do not.

Research to investigate jurors' sensitivity to pertinent variations in the substantive content of probabilistic information has only recently begun. Aschenbrenner, for example, (1978) found that irrelevant aspects of presentation influenced subjects' choices. A more recent study by Thompson (1984) was conducted to assess whether mock-jurors were sensitive to variations in the reliability of forensic tests which form the basis of probabilistic scientific evidence. Some subjects were informed that when hair sample analyses were conducted, the rate of false positive identifications was either 1% or 5%, while subjects in a control group received no information about the reliability of the forensic tests. The research results were consistent with those by Kahneman and Tversky (1973), i.e., there were no significant differences in the estimates of guilt by subjects who received different information about false positive rates. Thompson believes these results raise questions about the ability of jurors to evaluate scientific evidence of this type. A reasonable hypothesis is that some jurors are more competent at this task than others. Thompson reported only the mean estimates of guilt in each group. Mean scores are not diagnostic of sub-groups of jurors who may use the information differentially, i.e., whether some jurors use the information appropriately is unknown. Without additional information regarding jurors' individual mathematical abilities, it is impossible to determine whether there is a relationship between jurors' experience with and attitudes towards this sort of probabilistic information and their use of it in reaching a verdict.

The study by Faigman (1983) in which the frequency of matching blood type evidence was varied revealed that jurors adjusted their ratings of guilt in situations in which the frequency information was most pertinent (5%). These results showed some sensitivity to variations in the frequency of probabilistic information, albeit non-Bayesian. More information is needed about conditions in which jurors respond to these variations, and what factors predispose jurors to respond appropriately to the information.

Very few researchers have specifically examined jurors' reasoning errors in cases in which modification of the frequency probabilities is an independent variable. The studies by Faigman and Thompson are valuable in establishing that the introduction of probabilistic evidence into a decision-making task creates some unique problems not encountered in cases in which direct, eyewitness accounts comprise the evidence. These studies are further important in that they raise a number of questions about subjects' abilities to understand and use probabilistic information, and highlight issues that merit further investigation. Nonetheless, they leave little doubt that mock-jurors do not reason in accordance with the Bayesian model. Of course, this finding is not new to psychology (Barclay, Beach & Braithwaite, 1971).

Various applications of the Bayesian formula have been employed by cognitive scientists in studying human judgment to identify biases or systematic departures from an optimal standard (Fischhoff & Beyth-Marom, 1983). Hastie (1983) critically examined the application of Bayesian models to decision making, noting that the value of such normative models varies depending upon the extent to which the

following three criteria are present: (a) the normative model for the judgmental task is generally recognized as optimal; (b) the model can be mapped on to the task; (c) subjects represent the problem in the same frame as the experimenter. He concluded that in many instances, the Bayesian model is useful as a means to calculate performance limits, and as a mechanism to identify systematic biases, but that it is unacceptable as a framework for psychological theory (Rasinski, Crocker & Hastie, 1985). Findings such as those by Beach, Mitchell, Deaton and Prothero (1978), which illustrate that people do not reason in a Bayesian fashion, underscore the importance of this distinction.

F. Do jurors confuse the burden of proof with probabilistic evidence?

Variations in the decision-making task can result from a change in the substantive task-variables as well as a change in procedural task-variables, such as the response mode or the decision-rule. Procedurally, jurors' decision-making is guided to some extent by the judge's charge or instructions which incorporate a number of decision rules. For example, jurors are told whether their decision must be unanimous, or whether a majority decision will suffice. Hastie, Penrod and Pennington (1984) noted that the influence of this decision rule upon jurors' decision has been thoroughly investigated. A second vital decision rule is the applicable burden of proof which must be met before the moving party may prevail.

The law recognizes three separate and distinct burdens of persuasion or burdens of proof. The burden of proof is distinguished from the burden of production (the burden of going forward) which dictates who must first present evidence on an issue. The burden of

proof is the standard of proof or sufficiency of evidence that must be met before a conviction or finding of liability will be made. In criminal cases, the burden is "proof beyond a reasonable doubt." In civil cases, a lesser standard is applied, "proof by a preponderance of the evidence," often paraphrased as proof leading the factfinder to find the existence of a disputed fact more likely than its nonexistence. Loh (1985) has noted that the criminal standard focuses on the amount of doubt in the mind of the factfinder, whereas the civil standard focuses on the likelihood of the evidence. The third standard, "proof by clear and convincing evidence," is regarded as a higher threshold than the civil standard, but a lesser threshold than the criminal standard. This criterion applies in deportation cases, and some civil cases involving fraud.

In reviewing the judicial notion that jurors will confuse the burden of proof with probabilistic scientific evidence, Jaffee (1979) formulated a useful distinction between quality and quantity of proof. He noted that quantity of proof is what is at issue when the factfinder makes a decision, using either a preponderance of the evidence standard or the beyond a reasonable doubt standard. Probabilistic evidence, on the other hand, comprises a qualitative form of proof, just the same as eyewitness testimony is a qualitative form of evidence. When eyewitnesses testify, their credibility determines whether a rational or reasonable belief in the events attested to can be formed by the factfinder. When probabilistic evidence is presented, the issue is its relevance which affects the extent to which a rational or reasonable belief in the events can be formed by the factfinder.

The impact of the burden of proof instruction upon juror decision-making has received a fair amount of research attention since 1970 from Simon (1970), Thomas and Hogue (1976), and Nagel (1979). Renewed interest in the topic is apparent by the recent publication of two recent reports of studies of jurors' interpretations and applications of the burden of proof by Dane (1985) and Kagehiro and Stanton (1985).

One of the central issues to emerge when probabilistic evidence is admitted into evidence is whether jurors confuse this testimony with the burden of proof, particularly in a criminal trial in which the burden of proof is "beyond a reasonable doubt". This confusion may arise because jurors fail to distinguish between the quality of proof (direct or circumstantial evidence) and the quantum of proof (burden of proof) when the circumstantial evidence is probabilistic, because the qualitative evidence sounds quantitative. In other words, jurors may substitute the former for the latter.

In reviewing research on jurors' understanding of the criminal burden of proof, "beyond reasonable doubt," Dane (1985) observed that this concept may be one of the most confusing concepts used by the juror in his or her decision process. Earlier, jury researchers Wasserman and Robinson (1979) speculated that in civil cases it is particularly difficult to distinguish the task of discerning the facts from the task of applying the threshold probability for finding in favor of one of the parties; whereas in criminal cases, evaluation of the evidence and application of the burden of the proof are more readily discernible, thus less likely to be confused. On this basis, one would predict more confusion of the burden of proof and

probabilistic evidence in civil than in criminal cases. However, judicial concern over the potential for confusion has emerged only in criminal cases.

The purpose of much preliminary research on the influence of judicial instructions has been to determine whether jurors pay any attention to the instructions. With respect to burden of proof instructions, after reviewing taped mock-jury deliberations, Hastie, Penrod and Pennington (1984) found that subjects rarely referred to the burden of proof when reaching a verdict, a finding possibly attributable to jurors' failure to understand when to apply this standard of proof. (On the average, jurors total references to jury instructions during deliberations comprised only 25% of deliberation time.) Hastie et al. (1984) hypothesized that when a probabilistic issue was the focus of the trial, such as when identification of the perpetrator was in issue, the standard of proof would be more prominent in jury deliberations.

Prior research indicates not only that jurors have trouble understanding the burden of proof, but that they may not use the concepts as intended by the law: In a series of experiments conducted by Kagehiro and Stanton (1985), designed to assess the ability of subjects to differentiate between the three legal definitions of standards of proof, (preponderance of the evidence, clear and convincing evidence, and beyond a reasonable doubt), subjects were presented with a one-page written summary of an automobile accident, followed by pattern jury instruction defining the applicable burden of proof. The researchers reported that subjects, in this case,

psychology undergraduates, did not adjust their verdicts when applying different burdens of proof. In this between-subjects study, the legal definition had no effect on the dependent variables. In other words, the number of verdicts in favor of the plaintiff did not decrease as the standard of proof became stricter.

In a subsequent study by the same researchers, legal definitions from different jurisdictions were crossed with two different trial summaries to assess whether the legal language used to define the standard of proof in some jurisdictions would promote the application of the intended legal decision criteria better than others. The investigators found that definitions in existence vary widely in their ability to communicate to subjects the intended difficulty level. For example, definitions which are part of the federal pattern jury instructions are more likely to promote the use of the desired criteria than are pattern jury instructions recommended in California and Colorado. The researchers noted that subjects' performance was especially poor when they had nothing with which to compare the prescribed burden of proof. In actual trials, jurors are not given a set of two or three different burdens of proof to clarify the one which they must apply, i.e., no comparative definitions are provided. All jurors receive is a verbal definition of the burden of proof applicable to the facts they must decide.

One of the limitations of the Kagehiro and Stanton (1985) study is that it is uninformative about reasons for jurors' failure to distinguish between the burdens of proof. The direction of the error remains unknown: is the standard applied in civil cases too stringent?

Is the standard applied in criminal cases too lenient? Do jurors ignore the decision rule? Do different jurors develop juxtaposed definitions of the burden of proof that are not apparent when mean dichotomous dependent variables or mean ratings of culpability are reported? Early work by Simon and Mahan (1971) in quantifying the burden of proof showed that there was a tendency for jurors to apply too stringent a standard in civil cases (means of .75 or more were reported on a scale of zero to 1, whereas the mean estimate by judges was .55). That prosecutorial bias may be stronger in student mock-jurors than it is in real jurors because previous research by Zeisel and Diamond (1974) has shown that selection of college students rather than actual jurors in mock-jury studies results in more convictions than when real jurors serve as subjects.

However, one limitation of the Simon and Mahan (1971) survey is that their subjects (judges, jurors and students) were not given any particular legal definition to interpret, nor any fact pattern to which the definition was to be applied. Subjects were merely asked to provide a number between zero and one to express their understanding of the phrases "preponderance of the evidence" and "beyond a reasonable doubt." In actual court trials, jurors are not provided with both sets of definitions, and do not have the benefit of comparing the legal language of each. Nonetheless, Simon and Mahan found that both jurors and students rated the civil burden of proof as more than 70%, considerably higher than 51%, a common paraphrase of this threshold. In related research, Simon (1970) presented jurors with either a verbal definition or a quantitative definition of the burden of proof,

compared the number of convictions in the two groups and found them to be about the same. This between-subjects study has been criticized because there were no controls over subjective evaluations of the scalar values, i.e., there was no way to determine whether subjects who convicted using the verbal definition of beyond a reasonable doubt condition would also have convicted given a quantitative definition of the standard of proof.

An elaborate reseach investigation of jurors' verdicts using either the civil or the criminal burden of proof was conducted by a team of lawyers and psychologists, Sutton-Barbere, Teitelbaum and Johnson (1986). A videotape mock trial based on a fatal automobile accident was presented either as a civil wrongful death suit or as a criminal vehicular homicide case. After watching the trial, 72 different six-person juries deliberated to a verdict. Eighteen juries were hung. Thirteen juries who watched the civil trial returned verdicts finding the defendant liable; thirteen found the defendant not liable. Fourteen juries who saw the criminal trial found the defendant guilty; fourteen found the defendant not guilty. Jurors provided individual ratings of the burden of proof and of the likelihood that the defendant was liable/guilty. These data indicated that as many as 46% of the jurors assigned to the civil condition rated the preponderance of evidence threshold as between 70% and 100%, a more stringent standard than is intended by the law, thereby replicating the findings of Simon and Mahan (1971). In the criminal condition, jurors' ratings of the standard "beyond a reasonable doubt" indicated that a large number of jurors interpreted the criminal standard of proof far more leniently

than the law intends, i.e., 45% of the subject-jurors rated this standard as less than 90% likelihood. Thus, one explanation for the finding of apparent insensitivity to the burden of proof is that this standard is so ambiguous and poorly defined that jurors in both groups misinterpret it, and apply either a more stringent or a more lenient standard than the one they were instructed to use, thereby reassigning themselves to the other experimental condition.

Analyses by the same researchers of jurors' efforts to paraphrase the burden of proof have supported the notion that some jurors use a standard that is too harsh and that others use a standard that is too lenient. One plausible explanation for the failure by jurors to differentiate between the civil and the criminal burden of proof is that individual jurors apply different decision rules or models in evaluating the evidence against the burden of proof. For instance, some may apply a global top-down "goodness of fit" test, and others a bottom-up element-by-element mapping of evidence against the decision-rule. These trends have elements in common with two broad classes of jury decision-making strategies, i.e., verdict-driven versus evidence-driven decisions observed by Hastie, Penrod and Pennington (1984). Whatever the precise mechanism underlying these findings, the implications are clear: variations in the burden of proof alone are unlikely to result in a change in the perceived likelihood of guilt. Whether there is an interaction between the burden of proof and probabilistic evidence remains unknown.

In studying factors which influence decisions under conditions of uncertainty, cognitive scientists have attempted to distinguish task

effects from context effects (Payne, 1982). A decision-rule is typically classified as a task-related activity, and indeed, in the Kagehiro and Stanton study (1985), the burden of proof was treated as a task-related variable. In addition to this decision-rule, several other factors bear on the question of jurors' sensitivity to variations in the burden of proof. First, in reaching a verdict in either a civil or a criminal case, jurors are required to make a choice between two alternates: guilty/not guilty or liable/not liable. Tasks in which the decision involves the evaluation of two alternates often produce different results from those found when a judgment is sought, involving the same information. Psychologists have theorized that different strategies are employed by the decision-maker depending upon whether the elicited outcome is a choice or a judgment. For example, a choice between alternates is believed to stimulate alternate-based processing strategies such as "elimination by aspects" (Tversky, 1972); paraphrased as either a "pick one" versus a "reject one" strategy. Criminal cases typically present jurors with a choice: they can return a verdict of "guilty" or "not guilty". Civil cases typically involve not only a choice, but a determination of an appropriate amount to compensate a plaintiff, if jurors hold the defendant liable.

Psychologists have found that differences in the quality of the available options may also influence the strategy used by the decision-maker. For example, when one of the alternatives is not attractive, the "reject one" strategy is more common (Coombs, Donnell & Kirk, 1978). The attractiveness of available options in a jury trial may depend on the characterization of the case as either a civil or

criminal matter as much as the substance of the testimony before the jury. Thus, social context cannot be dismissed as a possible source of bias in the way in which jurors apply the different standards of proof to the same set of facts. The influence of social context and other situational factors which have been shown to affect decision-making are discussed in the next section.

The Influence of Situational Factors

In distinguishing task effects from context effects, Payne (1982) classifies structural characteristics of the decision-problem--such as the presentation-mode, the agenda, the constraints and the response mode--as task effects. By comparison, he designates as context effects those factors associated with the values and objects of the decision set, such as the overall attractiveness of the alternatives.

In the context of an actual trial as opposed to laboratory experiments, it is impossible to separate the burden of proof decision-rule from the broader social ramifications of the case, because the consequences of classification of a case as either a civil or criminal case are so markedly different. Criminal trials are associated with penalties and social opprobrium absent from civil trials. The entire juvenile diversion program is premised on sociological findings that in a criminal trial, the implications of labeling an individual first as a suspect, then as a criminal defendant, and ultimately as a convicted felon will negatively influence both the suspect and those who come to hear about the incident. One can reasonably hypothesize that jurors are not immune to such bias.

In the study by Kagehiro and Stanton (1985), all jurors were

presented with a civil trial summary, identifying the parties as plaintiff and defendant. The dispute in issue, an automobile accident, is more commonly the basis of a civil legal action than a criminal trial. The contextual cues in the experimental materials may have triggered different evaluative processes than would come into operation if indicia of a criminal trial were prominent. Conversely, perhaps minimal verbal cues such as the burden of proof itself which mentions "guilt"/"liability" function as adequate cues that the consequence of the decision is more or less socially punitive.

Thus, one feasible explanation for the finding that the conviction rate remains constant despite variations in the burden of proof is that jurors pay more attention to the nature of the offense than to the mechanics of the decision rule. Previous social science research has confirmed that social values or the social context can exert a powerful influence over juror decisions. For instance, in a study in which researchers systematically varied the penalty or consequence of the jury decision, they found that the number of convictions decreased as the severity of the penalties increased (Kerr, 1978). Increases in the severity of the consequences of the choice modified the conviction rate, while other independent measures, the subjects' ratings of guilt and likelihood of commission of the crime by the defendant remained constant.

One empirical study which provided support for the context-effect involved subjects' assessments of probabilistic statistical information. Frequency information about the number of crimes committed by two groups of people, one group being associated with more

serious crimes than the other, was presented to subjects (Rothbart, Fulero, Jensen, Howard & Birrell, 1978). Subjects were asked to make sample estimates of the frequency of crimes based on the data provided. Subjects assigned to the more serious crime condition produced frequency estimates that far exceeded those produced by subjects in the non-serious crime group, even though all subjects received the same frequency probability information. In other words, statistical judgments were influenced by factors irrelevant to the statistical issue, but which had strong negative social consequences. The perceived consequences of the subject matter alone triggered powerful inferential biases. Findings along these lines have led some psychologists to conclude that "strong equitable or prejudicial concerns that have little bearing on the probabilities of the evidence will affect the outcome by implicitly changing the standard of proof" (Wasserman & Robinson, 1979, p. 107).

Fischhoff and Lichtenstein (1976) noted that the finding that people's inferences and judgments are influenced by the attractiveness or costs and benefits of outcomes has been replicated in a number of experiments. This phenomenon has become known as the influence of the "problem frame" (Tversky & Kahneman, 1981). In general, people are more influenced by the threat of loss than the promise of gain. Thus, people are more likely to take action to avoid or minimize loss than to secure a comparable gain (Costanzo, Archer, Aronson & Pettigrew, 1986). When a problem-choice is framed in terms of the probability of winning, and is thus coded as a gain option, people may respond in a risk-averse manner, differently than when the same problem-choice is presented in a

way that emphasizes the probability of loss, which will probably be coded by the decision-maker as a loss option, triggering a risk-taking response. Tversky and Kahneman (1981) noted that framing outcomes in terms of overall wealth or welfare may attenuate one's emotional response to an occasional loss. This observation is pertinent to legal arguments presented by attorneys at the close of a trial. One attorney may frame an event as an uncompensated loss, while opposing counsel frames the same event as a cost incurred to achieve some benefit.

In a criminal trial, a finding that the defendant is guilty and will serve a jail-sentence can be framed as a gain for society, for instance, if the criminal is regarded as a public menace. Conversely, the same verdict may be framed negatively from the standpoint of an attractive defendant, who may be a first-time offender, likely to suffer social stigma and punishment unwarranted in the eyes of a jury, even if the technical elements of the commission of a crime have been proved. Similarly, a civil case which results in a finding of liability can be framed both positively and negatively, depending upon whether one considers the result from the vantage point of the plaintiff or the defendant. A verdict benefiting one party usually results in a cost to the opposing party, but the quality of the option set may vary enormously. For example, in a personal injury case in which the plaintiff is disabled, jurors may be more prone to return a verdict finding a wealthy defendant corporation liable than if the defendant is an unemployed friend of the plaintiff (Broeder, 1959).

Included in the set of situational demands which may exert some influence over a juror's decision is the type of response required by

the court. In criminal trials, the verdict is usually one of guilty versus not guilty, though jurors may be required to distinguish between different levels of criminal intent, and possibly to consider lesser included offenses. In civil cases, the preliminary response sought is a verdict that the defendant is either liable or not liable. In cases in which the plaintiff prevails, if damages are an issue, the jury must continue to deliberate to decide what amount of money to award to the plaintiff. To arrive at a total damage award, jurors may have to take into account a number of different types of damages, such as compensatory damages versus punitive damages. There is no literature on juror decision-making in determining an appropriate monetary award, and very few studies which have investigated juror decisions in civil cases. In laboratory studies, the option of a monetary award as opposed to the request for probability of gambling choice caused subjects to reverse their preferences (Lichtenstein & Slovic, 1971). Most juror decision-making research has focused on criminal cases, and less is known about the impact of options available in civil cases upon juror strategies.

Jury researchers have typically required mock-jurors to provide either judgment ratings (zero to 100) or guilty/not guilty verdicts, and sometimes both, as dependent measures. While the use of judgment scales has been criticized for lack of ecological validity, until recently there was not much concern that the modification of the response format might bias the decision-making process. Subject sensitivity to task demands raises fresh questions about generalizing from results in which subjects provide a judgment to situations in

which a subjects must choose between two alternatives. Payne (1982) argued that in situations of choice, a justification process usually follows (1982), whereas tasks which require a likelihood judgment on a scale of zero to 100 require no justification. He attributes inconsistencies in findings that come when the reponse format is varied to the justification process which accompanies a choice and which is absent in a judgment problem.

Another situational variable listed by Kaplan and Schersching (1980) as potentially important in an investigation of factors that bias juror decisions is time. A study by Wright (1974) in which time pressures in a complex decision-making choice task (buying a car) were varied showed that people's decisional strategies changed as time pressures increased. In situations of high pressure, people tended to focus on the probability of loss, i.e., on the negative aspects of certain options available, rather than on element-by-element mapping to select the best available option. Wright cautioned that this strategy might only be employed when the consequences of a final choice had some impact on the decision-maker, such as a personal investment in a car. Whether jurors who make decisions which have serious consequences for others will exhibit the same strategies is an open question. Mock-jury researchers have noted that subjects appear to take their task very seriously. Since data collection from real jury deliberations is prohibited, the impact of time pressures which jurors may impose on themselves, such as rushing through deliberations because of a desire to reach a verdict before dinner, cannot readily be assessed.

In assessing differences between various models of contingent

decision behavior, Payne (1982) isolated three dominant theories: cost/benefit, perceptual, and adaptive learning/production system models. In evaluating support for the latter model which posits that production systems account for the strategies and biases that decision-makers employ under conditions of uncertainty, Payne observed that individual differences in experience are likely to be strongly correlated with the degree and kind of task and context effects observed. Individual differences also comprise the third component of the Kaplan-Schershing information-integration model of factors which influence jury decisions. Research pertinent to the influence of individual differences as a source of juror bias is discussed next.

The Decision Maker

A. Scripts and Knowledge Schemata

A number of social and cognitive psychologists have demonstrated that the knowledge and experience which the decision-maker brings to the task may operate as a source of bias. For example, Rumelhart (1980), and Zadny and Gerard (1974) studied people's expectations, based on prior information or experience. Rumelhart (1980) posited that prior knowledge operates as a "script" or "schema", which affects the way in which new information is encoded, interpreted and retrieved from memory during a decision-making task. When new information is uncertain or ambiguous, reliance on pre-existing knowledge structures to interpret the events increases.

In cases in which probabilistic scientific evidence is introduced, jurors who have experience in working with statistical or mathematical

concepts may have an advantage over jurors who are inexperienced in this field. In a study of proportional reasoning strategies among the adult population, reported by Capon and Kuhn (1979), significant variability was found in the strategies and skills of 50 female shoppers who made decisions to purchase large or small packages of supermarket goods with different unit prices.

The role of prior knowledge, such as mathematical expertise, is under investigation by numerous cognitive researchers. For example, Shaklee and Hall (1983) reported that male and female undergraduates consistently used different strategies and varied in their accuracy when solving covariation problems. They theorized that these differences were attributable to subject differences in training and prior experience with mathematics.

The studies reported by Faigman (1983) included some analysis of the influence of individual differences upon verdict and use of probabilistic information. No significant differences attributable to subjects' education, mathematics background, age, gender, or prior statistical background were found. However, jurors with expertise or training in the field of statistics demonstrated significantly enhanced comprehension of the testimony by the statistician on the uses of Bayes' Theorem. These findings are not inconsistent with those of prior researchers who have reported that not only novices, but experts, too, make systematic errors in probabilistic decision tasks (Fischoff, Slovic & Lichtenstein, 1978; Bradley, 1981; Kahneman, Slovic & Tversky, 1982).

B. Attitudinal Biases

A second vital component of individual differences is attitude or affect. Interest by social and cognitive psychologists in affective as well as cognitive variables which influence judgments and decisions has prompted some research on ways in which jurors' attitudes may bias their judgments. Two jury simulation studies deserve mention in this regard.

First, a study conducted by Pryor, Taylor, Buchanan and Strawn (1980) is important because it examined not only the influence of jurors' attitudinal baises, but how those biases interacted with beliefs about circumstantial evidence. These researchers reported that most jurors believe that circumstantial evidence is not legal evidence, and that even the 40% of jurors who did acknowledge that circumstantial evidence was legal evidence believed that it should not be considered such. After jurors watched videotaped instructions from a judge explaining that circumstantial evidence should be taken into consideration in reaching a verdict, jurors' negative attitudes towards this evidence persisted. The researchers concluded that further judicial instructions and/or semantic or syntactic refinements to the exisitng instructions would not be effective in removing cognitive-affective biases of this nature, and that more extreme remedial measures were needed. The implications of this research are that the jurors may ignore or undervalue probabilistic scientific evidence not because it is probabilistic or scientific, but because it is circumstantial, rather than direct. The fact that the information is probabilistic and abstract may be merely a secondary factor in

discouraging jurors from according more weight to this evidence in reaching a verdict.

A second jury simulation study which investigated the impact of cognitive-affective biases was conducted by Casper, Benedict and Kelly (1985), who explored juror attitudes, cognitions and the hind-sight bias in civil search and seizure cases. They reported that the attitudes and beliefs of subjects (in this instance, undergraduates) strongly affected both their interpretation of the facts as well as the ultimate decisions reached.

Pryor et al. (1980) suggested that the question of whether the cognitive biases exert an influence over jurors' attitudes or whether jurors' attitudes precede cognitive biases was a matter which warranted further research. The study by Casper et al. (1985) attempted to answer this question by testing various path-analytic models of jury decision-making to specify the decison process. In the straight cognitive model, jurors' interpretations and reconstructions of the events were hypothesized to influence the size of the damage award. In a second model, jurors' attitudes were thought to influence the award, independent of the interpretation of the events. The influence of attitude was found to be statistically significant only for jurors with relatively constrained belief structures, i.e., jurors who exhibited highly consistent attitudes on a liberal-conservative ideology scale.

While ideological attitudes may have a bearing on how jurors respond to the substantive issues presented in the context of a search and seizure case, their influence in cases in which the substantive evidence under investigation is probabilistic testimony is unlikely.

In cases involving circumstantial testimony, other attitudes may be more pertinent. For example, Buchanan et al. (1978) predicted an interaction between juror biases and assessments of circumstantial evidence, hypothesizing that jurors who placed a high value on law and order would regard ambiguous circumstantial evidence as proof of guilt. Thus, judges who fear that jurors will accord too much weight to probabilistic circumstantial evidence have company, at least insofar as predictions about responses of certain jurors who regard the evidence as ambiguous are concerned.

One predisposition which seems worthy of examination in assessing factors which may influence the way in which jurors respond to probabilistic circumstantial evidence is attitude towards mathematics. Previous research has shown that mathematics confidence is significantly correlated with mathematics performance, and negatively correlated with mathematics anxiety (Clute, 1984). Just as students with high mathematics confidence have been shown to perform better on math tasks (Fennema & Sherman, 1977), one can hypothesize that jurors with high mathematics confidence will perform better in decision-making tasks in which mathematical concepts, such as probability, are prominent. While Faigman (1983) examined the influence of mathematics and statistical background (education) on comprehension of statistical concepts and ultimate verdict, he did not explore the influence of the subjects' attitudes towards mathematics or statistics, and this issue remains unaddressed.

Some researchers have investigated the interplay between self-report of mathematics proficiency and the costs and benefits of

different strategies available to solve a problem. For example, Christensen-Szalanski (1980) compared the strategies employed by students who rated themselves as mathematically non-proficient with those of business students, proficient in mathematics, in estimating company profits, given a specific set of information. The nonmathematical group was significantly less confident than the business group when using the available strategies. Psychological cost curves for the nonmathematical group also accelerated more sharply than did curves for the business group. The nonmathematical group invested more time when it was more costly to think and less time when it was less costly to think. Time constraints caused subjects to eliminate some strategies from consideration. Differences in mathematical aptitude affected the problem solver's confidence, accuracy and cost curve, but not the strategy selection process.

Costs and benefits of the decision outcome have also been shown to exert an influence on the decisionmaker. In examining the relationship between attitudes and probabilistic decision-making, Carr (1980) theorized that perceived costs and benefits of the outcome displace subjective likelihood away from rational probability. He classified subjects into one of two groups: "negative maximizers" and "negative minimizers." The classification into these groups was based upon consistent subjective likelihood reports by these subjects which were either significantly higher or significantly lower than the rational probabilities in a given circumstance. Using self-report and autonomic measures of anxiety, Carr found that subjects in the negative maximization group were significantly more anxious than subjects in the

negative minimization group. Carr studied whether the individuals consistently predicted highly positive or highly negative outcomes across different contexts and using different procedures. He found that classification as a minimizer or maximizer was an accurate predictor of biased responses so long as the outcome involved was negative rather than subjectively positive. In other words, when the outcome had positive personal consequences, jurors' subjective values entered into their judgments. In extending his general findings regarding decisions without personal consequences to jury decision-making, Carr speculated that jurors who are negative maximizers may focus on the negative effects of conviction for the defendant, and therefore argue in favor of acquittal. Conversely, a negative maximizer who focuses on the negative consequences of an acquittal for society may argue in favor of conviction. This hypothesis has elements in common with the framing theory of Kahneman and Tversky, but links that theory to an anxious attitudinal predisposition by suggesting that jurors most susceptible to the impact of a negative frame will be highly anxious jurors, and also that their probability estimates will vary most sharply from Bayesian norms.

Intervention Strategies to Minimize Juror Biases

One of the major findings to emerge from cognitive research into the influence of task-related variables is the pivotal role of the form of the information to be considered by the decision-maker. In trial, evidence is primarily oral (Sperlich, 1985), supplemented by a few tangible exhibits, if admissible under the rules of evidence.

Accordingly, some emphasis on the impact of variations in the verbal presentation of probabilistic testimony is warranted.

A. <u>The Influence of Linguistic Variables</u>

Research on the way in which people understand probabilistic language was conducted by Beyth-Marom (1982) who found that even among experts who regularly use terms such as "possible", "likely", "doubtful", "one can expect", there is a great deal of confusion, and a wide range of values associated with these terms. In other words, precisely what is intended and what is understood by the use of these terms is often ambiguous. When the probabilistic terms were applied to a particular set of facts as opposed to simple numerical scores, ambiguity increased. Some expressions, such as "doubtful" were reported to have a range of as many as 70 points on a scale of 100. Common terms with ranges as broad as 40 points included "likely", "one can assume", "small chance", and "chances are not great". In related research in which laypeople as opposed to experts served as subjects, Lichtenstein and Newman (1967) found similar confusion and ambiguity regarding terms of probability. However, Tversky and Kahneman (1983) reported that linguistic variables alone were not responsible for judgmental errors made by subjects who were presented with a series of inferential problems.

Linguistic variations more akin to those commonly used in expert trial testimony were examined by Fischhoff, Slovic and Lichtenstein (1977) to assess whether phrasing probabilities in terms of odds as opposed to percentages would deter subjects from severe overconfidence when making impossible or nearly-impossible predictions, such as

diagnosing the malignancy of ulcers or predicting the winners of 6-furlong horse races. Subjects showed no reluctance to use extreme odds, such as 1,000,000:1, and the odds formulation did not diminish overconfidence. When greater linguistic variation was introduced, some effect was noted in studies reported by Seaver, von Winterfeldt and Edwards (1978). In response to questions such as "What is the probability that in 1973, the population of Canada exceeded 25 million?" subjects were required to formulate a response either in percentages, odds, fractiles or log-odds. Responses phrased in percentages and odds were better than those in framed as fractiles. The researchers found that log-odds did not work at all well.

The impact of variations in the language in which probabilistic testimony is introduced was examined by Thompson (1984) by presenting mock-jurors with forensic testimony in which frequency probabilities were expressed either as a discrimination probability or as a percentage. When discrimination probabilities were used, subjects were more prone to make the prosecution error, and overweight the testimony. When simple percentages were used, subjects were more prone to make the defense error and ignore the evidence entirely. This study provides limited support for the judicial hypothesis that jurors may confuse the probabilistic evidence with the burden of proof, that is, providing the expert testifies using discrimination probabilities. The results are not helpful in determining whether the error is attributable to confusion with the burden of proof or some other bias in mock-jurors' statistical reasoning and decision-making when discrimination probabilities are considered. Whether jurors are susceptible to the

same bias in civil cases in which a preponderance evidence of proof applies remains an open question.

A number of educators and psychologists have hypothesized that the difficulty of mathematics problems is exacerbated by other linguistic variables, such as complex syntax or vocabulary. The biasing influence of complex linguistic variables was doubted by a group of mathematics researchers (Paul, Nibbelink & Hoover, 1986) who designed a study to isolate readability as the independent variable. They prepared 15 mathematical problems in prose form at three different readability levels, varying either vocabulary and syntax. The problems were administered to over 1000 children in grades 3 to 6. Variations in readability did not affect the students' ability to solve the problems, nor did it interact with grade, problem type or adjustment method (vocabulary or syntax). In sum, no effect of readability level on problem difficulty was found. If this finding were to generalize to adult jurors, one might hypothesize that modifications of linguistic variables alone would have little impact on jurors' abilities to cope with probabilistic scientific testimony.

Other commentators have questioned whether an oral presentation is effective in communicating unfamiliar or complex concepts to jurors. A few studies shed some light on this question. For example, there is evidence that pallid written information is remembered better than orally presented information, but that live presentations capture people's attention (Chaiken & Eagly, 1976). Findings by Petty and Cacioppo (1979) supported the first premise. They reported that once people are involved, written materials help to give recipients time to

consider the message arguments in detail, which is crucial to persuasion. Research along these lines was advanced by Myers, Hansen, Robson and McCann (1983) who tested the efficacy of cognitive intervention techniques. Subjects who received information about probability accompanied by Venn diagrams plus written verbal explanations were less affected by the presence or absence of key words and the inclusion of irrelevant information than were subjects who received no diagrams and less complete explanations. Myers et al. (1983) theorized that the role of explanation was crucial in promoting expertise among novices, and that explanation facilitated tranfer--in this case, the ability of the subjects to apply the information to story problems, not unlike written trial summaries used in mock-jury research). The experimental variables manipulated in this study are analogous to simplification devices employed by many expert witnesses who present probabilistic concepts to jurors.

An interesting experiment conducted specifically to discern whether explanations assist mock-jurors to understand and use probabilistic scientific evidence in reaching a verdict in criminal cases was undertaken by Thompson (1984). He presented subjects (psychology undergraduates) with a one-page factual summary of a bank robbery case based on United States v. Massey (1979), in which a bank teller observed that the robber had red hair beneath his ski-mask. Forensic expert testimony regarding the probability of matching hair samples taken from a ski-mask found in a suspect vehicle with that of any other individual by chance alone was expressed as a percentage, i.e., 2%. Some subjects received a full explanation of the implications to be

Some subjects received a full explanation of the implications to be drawn from this testimony. They were informed that differential weight should be accorded this evidence depending upon the size of the appropriate population. For example, in a population of 100,000, if the probability frequency is 2%, this means that 2,000 people could have hair matching the sample taken from the ski-mask. Other subjects received no explanation of the inferential weight of the testimony. Thompson reported significantly lower conviction rates (estimates of guilt) from subjects in the full implication condition. This result indicates that without a full explanation, many jurors may not understand the implications of the probabilistic testimony. Precisely what inference they draw remains unclear, as the only dependent variable reported was an estimate of guilt, not any other comprehension measures. The error rates reported are likely to be lower than those found when similar evidence is presented to real jurors, who may lack the exposure to quantitative concepts familiar to undergraduate psychology students who served as mock-jurors in the Thompson study.

C. The use of diagrams

In advising lawyers on how to organize and present statistical information, Finfrock and Spradlin (1978) point out that statistical information is usually very dull, and that numbers alone may not be comprehensible. They advise the use of percentages and ratios when explaining statistical data entered into evidence, though Thompson's (1984) work suggests that the former may be preferable. Finfrock and Spradlin also argue that exhibits used by the expert must be made available to jurors at the same time. In addition to written (verbal)

exhibits, pictorial or graphic exhibits may aid jurors in understanding the implications of the probabilistic testimony. Graphic presentations are recommended to catch the jurors' attention and make the testimony more vivid. Strawn and Munsterman (1979) particularly recommended that jurors be provided with visual aids to assist them in understanding complex evidence.

Some support for these recommendations can be found in the psychological literature. The few significant differences found to result from the presentation of pictorially illustrated versus non-pictorially illustrated information, reported by Taylor and Thompson (1982) occurred when (a) source credibility was crossed with mode of presentation, and (b) complexity of message was crossed with mode of presentation. However, the type of graphic illustration that will prove most effective has yet to be determined.

One study in which the form of the graphic presentation was varied was conducted by Wainer (1980). In this developmental study, graphic informations was displayed in the form of a bar graph, a line graph or a pie chart. Subjects' ability to answer questions about the information was influenced by the mode of presentation. For simple questions, bar graphs and pie charts produced the best results. For intermediate and comprehensive understanding, line graphs were superior. Thus, the effectiveness of visual aids in facilitating the application of information to a problem may vary. If these findings are applied to the courtroom, a question exists as to conditions in which juror comprehension may be facilitated by the use of graphic diagrams.

Related research indicates that the issue of the effectiveness of visual aids in obviating juror confusion bears more scrutiny. For example, investigations by Gick (1983) of conditions under which knowledge about one problem will transfer to an analogous problem in a different domain revealed that diagrams were not instrumental in facilitating spontaneous transfer, and that visual diagrams alone were not instrumental in facilitating the transfer of information from one set of facts to another. Expert testimony presented in a jury trial frequently requires jurors to apply findings from one set of circumstances to the facts presented by the case at hand.

D. The role of judicial instructions

While some commentators have focused attention on the mode of the presentation of the evidence, others have suggested that the instructions jurors receive about probabilistic testimony will play a crucial role in guiding jurors to use the evidence in an appropriate fashion, and recommendations have been made about jury instructions on the uses of this testimony. For example, Braun (1982) strongly advocates the role of the judiciary in instructing jurors on the proper use of probabilistic evidence, both regarding the probative value of the evidence (qualitative aspect) and how to weigh the evidence (quantitative aspect). To minimize the possibility that jurors will overuse scientific evidence, Pancerz (1983) recommended use of a limiting instruction to the effect that jurors must evaluate the probative value of the mathematical data in relation to all other evidence presented at trial.

More interesting remedial suggestions came from Buchanan et al.

(1978) who found pattern jury instructions on uses of circumstantial evidence to be of limited value. Instead, drawing on communication persuasion literature, they proposed that instructions should contain two-sided messages, and warnings to alert jurors to their biases regarding the weight of circumstantial evidence. In previous research, this inoculation technique has proved successful in overcoming subjects' biases. The application of these techniques to juror decision-making has not been empirically evaluated.

A third group of remedial intervention strategies concerns process guidance regarding the decision to be made by jurors. For example, Strawn & Munsterman (1979) suggested that jurors be given guidance on how to deliberate. Advice along these lines has emerged in the debate over factors that promote the "correct" use of probabilistic information. For example, in discussing people's tendency to ignore base-rate information, Beyth-Marom and Fischoff (1983) noted:

> The first step in helping people to improve judgment is helping them organize their thinking. If information is selected and organized for them, they generally show a qualitiative understanding of diagnosticity. (p. 1194.)

To assist jurors in thinking about probabilistic information, Finkelstein and Fairley (1970) recommended that jurors be instructed in the use of Bayes' Theorem to modify their initial subjective estimates of the probability of the guilt of the defendant. For example, suppose an expert testified that the odds were 1000:1 that a fiber found on the victim's clothing would match fibers taken from the defendant's carpet. If a juror's subjective estimate of guilt based on other qualitative evidence were .10, by applying Bayes' conditional probabilities

formula, the likelihood of guilt would be computed as .99.

In an interesting cognitive intervention study by Faigman (1983), the recommendation of Finkelstein and Fairley (1970) was tested by providing mock-jurors with a Bayesian explanation of evidence to assist them in evaluating probabilistic testimony concerning matching blood types. An expert witness, a statistician, explained to jurors how to apply Bayes' Theorem to the blood type frequency evidence introduced by another expert witness, a physician. Unfortunately, Faigman's experimental design did not incorporate a control group that was not exposed to the Bayesian presentation, so the precise effect of this technique was not assessed. Nonetheless, Faigman reported that subjects underused the probabilistic evidence in comparison with the Bayesian model.

Loh (1985) cautioned that the introduction of decision-making technology such as Bayes' Theorem into jury deliberations may dehumanize justice or upset the delicate equilibrium of the legal system, just as rendering the jury instructions more comprehensible may result in fewer convictions (Severance, Greene & Loftus, 1984). Other objections to the introduction of Bayes' Theorem include the fact that it often does not work (Williams, 1983); that it will provide a spuriously exact probability (Loh, 1979); that it will cause quantifiable evidence to become the focus of the trial (Jonakait, 1983); and that it makes no provision for instances in which impeachment testimony is introduced or in which there is insufficient evidence of the crime charged (Brilmayer & Kornhauser, 1978). Consequently, this cognitive intervention technique is unlikely to gain

much support in the legal community.

Summary

The foregoing review examined judicial hypotheses about juror competency to decide cases involving probabilistic, circumstantial evidence in light of social science research that bears on this issue. Four broad areas of research were examined: contingent decision-making considering both task effects and context effects; individual differences and cognitive intervention techniques.

On the whole, jurors appear to underuse rather than to overuse probabilistic information, but there is some evidence that a small percentage of subjects may overweight probabilistic testimony. Several questions follow from the preceding research, some of which are addressed in the present research. For example, Faigman (1983) wondered whether there is an interaction between the causal relevance and an extreme degree of probabilistic evidence. Little is known about the impact of variations in the magnitude of frequency probability evidence. This issue is examined in Study One. One common conclusion by psychological researchers is that courts should be more concerned with impressing upon jurors the relevance of probabilistic techniques and less concerned that jurors will be overwhelmed by their complexity. In Study Two, the influence of variations in the form of the presentation of probabilistic evidence are examined. Study Two also explores the relationship between the burden of proof and mock-jurors' uses of probabilistic evidence.

CHAPTER THREE

VARYING INCRIMINATING FREQUENCY PROBABILITIES IN CRIMINAL CASES

Study One

Purpose and Rationale

In response to heightened interest in the debate over problems jurors may experience with quantitative evidence and concepts, some preliminary research has examined the way in which mock-jurors use this information in reaching a verdict. The studies by Faigman (1983) and Thompson (1984, 1985) in which frequency probabilities were varied, left several matters unresolved, including the question of the inferences which jurors draw from such evidence. For example, Jonakait (1983) posed the question as to whether jurors understand that frequency probability evidence merely helps to define the class of suspects, and that its contribution to establish proof beyond a reasonable doubt is small. He wondered what goes on in the minds of jurors who hear that there is one chance in 1000 that incriminating evidence comes from a source other than the defendant. To date, psychologists have not addressed these questions and have gained relatively little insight into mock-jurors' comprehension and interpretation of frequency probabilities.

One way to assess jurors' inferences when probabilistic evidence is presented is to examine the fact-finding process in more detail. In law, Brown (1943) and Austin (1984) have noted that it is common to distinguish between at least two levels of factual findings: evidentiary facts and ultimate facts. Evidentiary facts are inferences based on testimony regarding what the witnesses saw, heard, etc., and

consist of the conclusions drawn by the factfinder concerning this testimony. In other words, an evidentiary fact is what the factfinder makes of the evidence or determines to be the most likely occurrence, given a series of controverted facts. Ultimate facts, by comparison, are legally cognizable facts, i.e., legal conclusions once the law has been applied to the evidentiary facts. For example, an evidentiary determination may be that the defendant killed his companion during a fight, while the ultimate legal fact is that the defendant is not guilty of murder because the act was committed without the necessary culpable mens rea, e.g., accidentally, or in self-defense. The focus of much prior jury research has been on the determination by mock-jurors of ultimate or legal facts. Only more recently have psychologists begun to investigate components of the decision-making processes, such as the relative weight accorded to testimonial evidence that is evaluated by jurors in the process of determining the legal facts. Studies along these lines include those by Devine and Ostrom (1985), Casper et al. (1985), Rasinski, Crocker and Hastie (1985), and Kaplan and Schersching (1980).

To assess whether jurors misuse probabilistic evidence, more information is needed about subjects' preliminary inferences regarding the evidentiary facts, i.e., what jurors make of the probabilistic evidence. Moreover, in view of widespread speculation that jurors confuse the burden of proof and probabilistic evidence, more information is needed regarding the manner in which jurors apply the decision-rule or burden of proof to evidentiary facts which are probabilistic.

The discussion in Chapter Two of applicable case law revealed that in most trials in which probabilistic frequency evidence is introduced, the controversy increases when the probability of a random match is fairly low, for instance, when it is less than 1%. There is less concern that jurors will misuse statistical evidence when the probability of a random match is high, for instance, when it is over 20%. The variations in the frequency probabilities introduced in the Faigman (1983) study were high numbers: 40/100, 20/100 and 5/100, arguably not very probative of guilt in a criminal case, although in a civil case, frequencies in this range may be more common. Cast in this light, the Faigman findings that mock-jurors tended to ignore the statistical evidence is less surprising. Subjects may have inferred quite reasonably that discrimination probabilities of 40/100 or 20/100, while somewhat relevant, did not merit a great deal of weight. The independent variables in the Faigman study, therefore, are of limited value in determining whether subjects tend to discount or ignore probative scientific evidence.

While the Faigman study did not speak well to the question of underuse of scientific evidence, it was instrumental in distinguishing between the subjects' evidentiary factual inferences without the probabilistic testimony, their inferences following presentation of the probabilistic testimony and their ultimate factual inferences regarding the defendant's guilt. Specifically, subjects were asked to estimate the likelihood that blood samples found at the scene of the crime were the defendant's, and also the likelihood that the defendant committed the crime charged, before rendering an ultimate verdict. Thus, some

insight into mock-jurors' comprehension of the probabilistic evidence, distinct from the role it played in determining their verdict, was gained.

Some caution is warranted in evaluating the Faigman (1983) findings because the experimental procedures used to assess subjects' inferences and decisions may have had a biasing effect on the results. At either one, two or three different points during the experiment, subjects provided estimates of (a) the likelihood that the blood found was that of the defendant and (b) the likelihood that the defendant committed burglary. Subjects each provided one, two or three "ultimate" verdicts on the legal issue to be determined: (a) prior to hearing the probabilistic evidence; (b) after reading the probabilistic evidence; (c) after reading a Bayesian explanation of the weight of the evidence. Faigman found that the number of probes significantly affected the mock-jurors' likelihood estimates: subjects who provided three sets of estimates produced significantly higher estimates that the matching blood came from the defendant than did subjects who provided only two sets of estimates. However, no significant differences were found for subjects who received two probes versus one probe. By way of explanation, Faigman noted that subjects who explicitly stated a prior probability of guilt before hearing the probabilistic evidence may have been sensitized to the blood-type evidence, and went on to add that the increased attention to the statistical evidence was more likely responsible for the difference than quantification of the non-statistical evidence.

By comparison, in the Thompson (1984, 1985) studies, more

incriminating frequency probabilities were presented (2% or 1% incidence of suspect characteristic in a target population), increasing the value of this research in assessing whether mock-jurors tend to underuse or overuse probabilistic evidence. In the first study, in which a 2% frequency probability was used, a 13% incidence of the prosecutor's fallacy was reported, i.e., 2% of the subjects assumed this meant there was a 2% likelihood that the defendant was innocent, and estimated that the likelihood that the defendant was guilty was 98%. In the same study, Thompson reported a 12% incidence of the defense fallacy, i.e., 13% of the subjects failed to revise their estimates of the defendant's guilt after receiving the incriminating frequency information. In a second study by Thompson (1985), in which a slightly more incriminating frequency probability was presented, (1% incidence rate of the suspect characteristic in the population) subjects read summaries of arguments presenting either the prosecutor's or the defense fallacy. Accordingly, higher incidence rates for both judgmental fallacies (28% and 68%, respectively) were reported. Modifications in the experimental materials and procedures used from one study to the next (i.e., the presentation of summaries of fallacious arguments) prevent detection of the influence of variations in the frequency probabilities upon the incidence of either the prosecutor's or the defense fallacy.

One additional problem with both Thompson (1984, 1985) studies is that there is no baseline measure, such as might be provided by a control group; i.e., there is no way to assess what impact the probabilistic evidence had upon mock-jurors' decisions because there

was no condition in which no frequency evidence was presented, and no measure of defendant's guilt in the absence of the target probabilistic evidence. Moreover, no comprehension measures were reported. Thus, while Thompson's studies assess some judicial hypotheses about uses and misuses of scientific evidence, the dependent measures were not diagnostic of mock-jurors' inferential processes.

There is no prior research on the relationship between frequency probability evidence and the burden of proof, a second area of inquiry which merits further attention. In the Thompson (1984, 1985) studies, for example, subjects were not informed of the applicable burden of proof, and simply made an estimate of the defendant's guilt based on uncontroverted information. In the Faigman (1983) study, subjects were presented with controverted facts, and were instructed to return a guilty verdict only if the evidence warranted this beyond a reasonable doubt. In neither case were instructions on the meaning of reasonable doubt provided. One reason it may be fruitful to include an instruction on reasonable doubt when assessing the inferences which jurors draw from probabilistic evidence is that this instruction may assume prominence when jurors evaluate quantitative and non-quantitative evidence in reaching a verdict. The probabilistic nature of the evidence may cause jurors to scrutinize all evidence more closely. Jurors may be reluctant to return a conviction on the basis of probabilistic evidence. Emphasis on the burden of proof instruction will facilitate an examination of the interaction between this task variable and jurors' inferences about probabilsitic evidence. While Faigman's data permitted a comparison between estimates of defendant's guilt and ultimate verdict, he never actually assessed this relationship. Subjects whose estimates of guilt were identical may have returned different verdicts, depending upon the interpretation they accorded to the criminal burden of proof. Neither the Faigman nor the Thompson studies addressed the judicial concern that probabilistic evidence will be confused with the burden of proof.

The purpose of the present study was to gather some preliminary information about the way in which mock-jurors comprehend and apply probabilistic information presented in the context of a criminal case when probative incriminating frequency probability evidence is systematically varied. Particular questions of interest were:

1. What inferences do mock-jurors draw from probabilistic information? Do they compute the number of people in a suspect population who possess the same characteristics as the defendant based on the probabilistic evidence? Do they convert the frequency probability into a probability estimate of defendant's guilt? If inferential errors occur, which are most common?
2. Do mock-jurors accord more weight to the probabilistic evidence when the probability of a random match is lower?
3. When probabilistic evidence increases or decreases in incriminating value, does this variation influence the way in which other evidentiary facts are perceived and weighed?
4. Do mock-jurors confuse the burden of proof with frequency probability evidence? Are mock-jurors equally likely to confuse the probabilistic information with the burden of proof when the probative value of the evidence is varied?
5. Is there a relationship between susceptibility to the prosecutor's fallacy and mock-jurors' interpretations of the criminal burden of proof?

Hypotheses. A major hypothesis was that subjects assigned to experimental conditions in which the frequency probability of a random match with the suspect blood type was lower (and in which the blood type was therefore more rare), would attribute more weight to the statistical evidence than would subjects in experimental conditions in which the frequency probability of a random match was higher. Accordingly, it was predicted that the estimates of the defendant's guilt and the conviction rate in the former groups would exceed those in the latter groups.

Second, it was hypothesized that instances of the prosecutor's fallacy would be more frequent when the probability of a random match was lower, i.e., in which the quantitative evidence was more extreme (5% or less). Third, it was hypothesized that jurors whose interpretation of the burden of proof was the complement of the frequency probability evidence would be more likely to confuse the frequency probabilities with the burden of proof. For example, suppose a juror interpreted beyond a reasonable doubt to mean that a confidence level of 95% regarding the defendant's guilt had to be met in order to return a verdict of guilty. It was hypothesized that confusion of the burden of proof would be more likely if that juror were assigned to the third experimental group in which the frequency probability is the complement of that number, i.e., 5%, than if that juror were assigned to another experimental group in which a different frequency probability was supplied.

Method

Subjects. Subjects were 233 psychology undergraduates attending the University of Washington who received credit for completing the questionnaire. More females (61%) than males (39%) participated in the study.

Materials. Experimental materials consisted of a one-page narrative summary of 20 evidentiary facts presented in a homicide case, based on an actual trial over the murder of a businessman in his apartment the day before an important business meeting with his partner of many years. After the murder, the surviving business partner had deep scratches on his face which he claimed were received in a minor bicycle accident during the preceding week, although no witness testified in corroboration of his alibi. The grand jury indicted the surviving partner for murder. Forensic experts gathered blood samples from the victim, his apartment floor, and from the suspect. The blood type of the victim was O, while that recovered from his apartment floor was type A2B. The surviving partner submitted to a court-ordered blood-test which revealed that he had type A2B blood. Population data revealed that blood type A2B occurred in a specified percentage of the population of Boise where the murder took place. The suspect was tried for the murder of his partner.

Following the one-page factual summary, subjects were given an explanation of the criminal burden of proof, using the standard Washington Pattern Instruction for Criminal Cases:

> A reasonable doubt is one for which a reason exists. A reasonable doubt is such a doubt as would exist in the mind of a reasonable person after fully, fairly and carefully considering all of the evidence or lack of evidence. If, after such considerations, you have an abiding belief in the truth of the charge, you are satisfied beyond a reasonable doubt.

Dependent Measures. A written questionnaire was designed to gather five types of information:

1. Verdict, expressed verbally as "guilty" or "not guilty" and also as a probability estimate on a scale from zero to one hundred (1 = not guilty, 100 = definitely guilty). In addition, subjects provided a quantitative estimate of the extent of their doubt that the defendant was guilty, on a scale from zero to 100 (0 = no doubt, 100 = definite doubt).
2. Weight accorded to twenty evidentiary facts in determining the guilt of the accused, expressed on a scale from -100 to 100 (negative values -100 to -1 indicated "not guilty"; positive values 1 to 100 indicated guilt). Sample facts listed for evaluation included: "Greg had scratches on his face on September 23rd, 1985" and "Bill's blood type differed from samples of blood found in the hallway of his apartment."
3. Identification of evidentiary facts considered in reaching a "guilty" verdict, and considered in reaching a verdict of "not guilty." This was achieved by asking subjects to divide 100 points among the twenty evidentiary facts listed previously. First, subjects assigned 100 points to facts indicative of defendant's guilt; second, they assigned another 100 points to

facts indicating that the defendant was not guilty. There was no requirement to assign points to all 20 facts listed.

4. Factual inferences drawn from the target frequency probability evidence. Subjects were asked to state (a) how many persons in the population possessed the suspect blood type, and (b) the odds that the blood samples taken from the victim's apartment belonged to the defendant. In both instances the correct response was to be selected from among 10 forced-choice alternatives.
5. Ratings of the perceived reliability of (a) eyewitness testimony, (b) statistical evidence and (c) blood tests on a scale from zero to 100 (zero = totally unreliable, 100 = very reliable).

Procedures

Design. A 1x5 design was used in which the independent variable was the probability of a random match between the defendant's blood type and the suspect blood type in a given population of 100,000. There were five levels to this variable: four experimental groups, in which the frequency probabilities were .1%, 1%, 5%, and 10%, respectively, and a fifth control group in which no frequency probability information was presented. The experimental manipulation consisted of varying the final sentences listing the fats in evidence. In the control condition, the final paragraph read:

> The analysis of blood samples reveals that Bill Taylor has type O blood. The court orders Greg Johnson to submit to a blood test, and he is found to have blood type A2B. While it is clear that the samples of blood taken from the apartment hallway are not type O, their precise identity is inconclusive.

In Groups One, Two, Three and Four this paragraph was modified to indicate that the defendant's blood type matched that found at the scene of the crime, and also to provide subjects with the target frequency probability information. For example, in Group Four, in which the probability of a random match between the defendant's blood type and that found at the scene of the crime was 10%, the final paragraph was:

> The analysis of blood samples reveals that Bill Taylor has type O blood. The court orders Greg Johnson to submit to a blood test, and he is found to have blood type A2B. The sample of blood taken from the apartment hallway is also type A2B. Population data reveal that blood type A2B occurs in ten per cent (10%) of the population of Boise, Idaho.

Subjects were informed that there were 100,000 people in Boise, Idaho; accordingly, they could calculate the number of persons in Boise who possessed the suspect blood type. The relationship between the independent variables and the corresponding number of persons in the suspect group is presented in Table 1. Subjects read written factual summaries attached to a pencil-and-paper questionnaire, and worked individually. The subjects spent between 45 minutes and one hour working through the materials.

Table 1

Study One: Independent Variables and Corresponding Size of Suspect Pool

	Independent Variable	Inference
Group	Frequency Probability	Size of Suspect Pool
1	One tenth of one per cent (.1%)	100 people
2	One per cent (1%)	1,000 people
3	Five per cent (5%)	5,000 people
4	Ten per cent (10%)	10,000 people
5	No conclusive match between blood of defendant and blood in victim's apartment	unspecified

Results

Ultimate Verdict. A majority of mock-jurors in each experimental group found the suspect not guilty of murdering his business partner. One-way analysis of variance revealed a significant main effect, F (4, 228) = 5.81, $p < .01$. These results are presented in Table 2. Overall, 23% of the subjects returned a guilty verdict and 77% of the subjects acquitted the defendant. Group means are displayed with other descriptive analyses in Table 5.

Table 2

One-Way Analysis of Variance of Student Mock-Jurors' Verdicts

Source	SS	*df*	MS	*F*	*p*
Groups	3.79	4	.95	5.81	.0002
Error	37.15	228	.16		
Total	40.94	232			

Post-hoc contrast tests indicated that the means in Groups One, Two and Three differed significantly from the mean in the control group, However, differences between the means in Groups Four and Five were not significant, t (228) = -3.39 (Group One), -3.11 (Group Two), -4.13 (Group Three), and -1.18 (Group Four); p < .01. See Table 3.

Table 3

A Posteriori Contrast Tests Between Group Means for Verdict

Contrast	Value	S. Error	*df*	*T* Value	*T* prob.
1, 5	-.27	.08	228	-3.39	.001
2, 5	-.26	.08	228	-3.11	.002
3, 5	-.34	.08	228	-4.13	.000
4, 5	-.1	.08	228	-1.18	.238
5, other groups combined	-.24	.06	228	-3.81	.000

Mock-jurors' probability estimates of the defendant's guilt are shown in Table 5. Univariate analysis of variance revealed a significant main effect for group, F (4, 228) = 4.67, $p < .01$. These results are displayed in Table 4. The overall mean estimate of the defendant's guilt was 39%. Chi-square analysis of group by verdict results showed a significant difference between the observed and expected number of convictions in each group. Chi-square (1) = 21.58, $p < .01$. Differences between the mean conviction rate, the mean estimates of defendant's guilt, and mean doubt of defendant's guilt by group, are displayed in Table 5.

This pattern of results shows subjects' numerical estimates of the defendant's guilt conformed to the hypothesized rank order, with the highest mean estimate of guilt in the group in which the frequency probability was lowest (1/1000), and steadily diminishing estimates in

Table 4

One-Way Analyses of Variance of Student Mock-Jurors' Estimates of Defendant's Guilt

Source	SS	df	MS	F	p
Group	101619.78	4	2654.95	4.67	.0012
Error	129498.87	228	567.97		
Total	140118.65	232			

Table 5

Mean Conviction Rate (Percentages), Mean Probability Estimates of Defendant's Guilt and Mean Doubt of Guilt by Group

	Incriminating Frequency Evidence					
	.001	.01	.05	.1	control	Grand Mean
Percent guilty verdicts	31	30	38	14	4	23
Mean estimate of guilt	47	45	40	34	29	39
Mean estimate of doubt	50	58	58	61	70	59
N =	48	44	45	44	52	233

Groups Two, Three and Four. The lowest mean estimate of guilt occurred in the Group Five in which no frequency probability was presented. A posteriori tests contrasting means in Groups One, Two, Three and Four with the mean in the control group were significant for the first three groups (p < .01), but not for Group Four. See Table 6.

Comparisons within each group of mock-jurors' verdicts with the estimates of defendant's guilt revealed that subjects convicted the defendant using an estimate of guilt ranging between 50% and 99%. In Group One, for example, in which 31% of the subjects voted to convict the defendant, only three subjects (2%) returned estimates of

Table 6

A Posteriori Contrast Tests Between Group Means for Estimates of Defendant's Guilt

Contrast	Value	S. Error	*df*	*T* Value	*T* prob.
1, 5	18.28	4.77	228	3.83	.000
2, 5	15.67	4.88	228	3.21	.002
3, 5	11.53	4.85	228	2.38	.018
4, 5	6.26	4.88	228	1.28	.201
5, other groups combined	12.94	3.75	228	3.45	.001

defendant's guilt of 90% or more. The remaining 29% of the convictions in this group came from subjects whose estimates of guilt ranged between 62% and 90%. In Group Two, the lowest estimate of guilt producing a conviction was 65%; the highest was 90%. In Group Three, in which 38% of the subjects convicted the defendant, only three subjects (4%) believed there was a 90% probability or more that the defendant was guilty. The remaining 31% of the convictions in Group Three came from subjects whose estimates of guilt ranged between 50% and 90%. In Group Four, 4% of the subjects concluded that the defendant was 90% guilty, and the remaining 11% of the convictions came from subjects who believed the defendant was 60% to 90% guilty. Thus, the higher conviction rate found in Group Three is partially attributable to the fact that a number of subjects in this group

interpreted the burden of proof somewhat less stringently than did subjects in the other experimental groups. In other words, these subjects employed an unusually low criterion of "beyond a reasonable doubt."

In sum, as shown in Table 7, while verdict measures produced an overall conviction rate of 23%, subjects' numerical estimates of guilt indicated that as few as 2% of the subjects found the defendant more than 93% guilty. Only 5% of the subjects found the defendant more than 90% guilty.

Table 7

Overall Percentage and Frequency of Subjects who Convicted Using an Estimate of Guilt Below or Above Ninety Percent

	Verdict					
	Guilty		Not Guilty			
Likelihood of guilt	Percent	f	Percent	f	Percent	Total
0 - 89%	20	44	80	178	95	222
90 - 100%	82	9	18	2	5	11
Percent Total:	23	53	77	180	100	233

To obtain another measure of the influence of variations in the probabilistic evidence upon subject's decisions regarding the defendant's guilt, subjects were asked how much they doubted the

defendant's guilt. Doubt was affected by the probabilistic evidence, just as were estimates of guilt. One-way analysis of variance revealed a significant main effect for group in response to this question, F (4, 228) = 3.81, $p < .01$. These results are presented in Table 8. Subjects in the control group expressed the most doubt regarding guilt (M = 70%), while subjects in the group with the most incriminating frequency evidence expressed the least doubt (M = 50%). The difference in mean ratings by subjects in intermediate groups Two, Three and Four, with frequency probabilities of 1%, 5% and 10%, was less marked (Ms = 58%, 58% and 61%, respectively). These results are included in Table 5.

Table 8

One-Way Analysis of Variance of Student Mock-Jurors' Doubt of Defendant's Guilt

Source	SS	df	MS	F	p
Group	9915.96	4	2478.99	3.81	.005
Error	148519.79	228	651.40		
Total	158435.75	232			

A posteriori tests contrasting means in Groups One, Two, Three and Four with the mean in the control group were significant for the first three groups, but not for Group Four. The mean in the control group was

significantly different from the mean of all four experimental groups combined. These analyses are displayed in Table 9.

Table 9

A Posteriori Contrast Tests Between Group Means for Doubt of Defendant's Guilt

Contrast	Value	S. Error	*df*	*T* Value	*T* prob.
1, 5	-.27	.08	228	-3.39	.001
2, 5	-.26	.08	228	-3.11	.002
3, 5	-.34	.08	228	-4.13	.000
4, 5	-.1	.08	228	-1.18	.238
5, other groups	-.24	.06	228	-3.81	.000

To determine the extent to which estimates of defendant's guilt and amount of doubt that defendant committed the crime were congruent, subjects' response scores for the two questions were summed. If a subject rated the defendant as 40% guilty, and was then asked how much he or she doubted that the defendant committed the crime, on a scale of zero to 100, one might expect a complementary response to the latter question, such as 60%. In fact, the sum of as many as 40% of the subjects' scores exceeded 100. The sum of 22% of the scores exceeded 120. These findings indicate that a fair number of the subjects may have misunderstood the question regarding the extent to which they

doubted that the defendant committed the crime. Note, however, that these subjects were not instructed that their answers to the two questions should add to 100. Alternative causes of this finding are that subjects' addition is poor, that their reasoning is flawed, or that they have an ambivalent response to the evidence.

Another way to assess mock-jurors' performance is to compare their responses with those of "a rational Bayesian" given the same information. Bayes' Theorem[1] facilitates an assessment of the influence of probabilistic evidence upon mock-jurors' decisions providing there is an available baseline measure of mock-jurors' estimates that the defendant is guilty (Finkelstein & Fairley, 1970), exclusive of the frequency probability evidence. In this case, the _a priori_ estimate of guilt was provided by mock-jurors in the control group who rendered a verdict in the absence of frequency probability information. Mock-jurors in the control group provided both a probability estimate regarding the defendant's guilt ($\underline{M}$ = 29%) and an ultimate verdict (mean conviction rate of 4%). Accordingly, the estimates of guilt and conviction rates in each of the four experimental groups were compared with similar predictions using a Bayesian analysis.

For example, suppose that prior to hearing the results of the blood tests, a mock-juror in Group Three has a subjective probability of .29

---

[1] The following expression of Bayes' Theorem was selected:

$$p(A/B) = \frac{p(B/A)\ p(A)}{p(B/A)\ p(A) + p(B/\text{not } A)\ p(\text{not } A)}$$

that the defendant murdered his partner, i.e., $p(A) = 0.29$. Therefore, the probability that the defendant did not murder his partner, based on the prior evidence, is .71, i.e., $p(\text{not } A) = 0.71$. The likelihood of a random match between the blood type of the defendant and the samples taken from the victim's apartment is .05, i.e., $p(B/\text{not } A) = 0.05$. The probability of a match between the blood type of the defendant and the samples taken from the victim's apartment in the event that the defendant did murder his partner, and in the course of a struggle, became injured and bled on the floor, is assumed to be 1, i.e., $p(B/A) = 1.0$. In substituting these values into the equation, it is clear that mock-jurors who held a subjective belief that the defendant was 29% guilty before taking into consideration the probability evidence, should revise their estimates upwards in light of the scientific evidence, to conclude that the likelihood that the defendant is guilty is 89%, as follows:

$$p(A/B) = \frac{(1.0)\ (0.29)}{(1.0)\ (0.29) + (0.05)\ (0.71)} = 0.89$$

Depending upon the mock-jurors' interpretation of the burden of proof, this estimate may or may not be considered adequate for conviction. Similar comparisons can be made regarding mock-jurors' verbal verdicts, using the percentage of guilty verdicts of the control group, to provide an _a priori_ conviction-rate. Of course, these comparisons are subject to greater criticism because of lack of uniformity in the application of the burden of proof to the evidentiary

facts. Differences between the actual responses of the mock-jurors in each group and the estimates of a hypothetical Bayesian rational person, given the same information, are illustrated in Figures 1 and 2.

<u>Factual inferences based on probabilistic information</u>. To diagnose causes of the predicted underuse of the evidence, subjects were asked to state, based on the evidence presented to them, how many persons in the Boise population were included in the suspect pool possessing the same suspect blood type as the defendant. To select the correct answer to this question, subjects had to recall the size of the population (given as 100,000), and to recall the given percentage of the population having the same blood type as the suspect, and then multiply the two numbers. For example, in the group in which subjects were told that 1% of the population had the same blood type, they had to multiply 100,000 by 1/100 to determine that the appropriate answer was 1000. Then, subjects had to select the correct response from a series of 10 forced-choice alternatives (10; 50; 100; 500; 1,000; 5,000; 10,000; 50,000; 100,000; I cannot say how many). Thirty-four percent of the subjects overall did not answer this simple arithmetic question correctly. The error rate was highest in Group One in which the frequency probability was most probative, i.e., had the smallest number (1/1000). The majority of the errors by subjects in Groups One, Two and Four came from subjects who selected an incorrect multiple of ten in response to this question. The lowest error rate came from subjects in Group Three, in which the frequency probability evidence was not a multiple of ten (i.e., there was a 5% probability of a random match).

Figure 1. Student Mock-Jurors' Estimates of Defendant's Guilt as a Function of Frequency Probability Evidence versus "Rational" Bayesian Estimates.

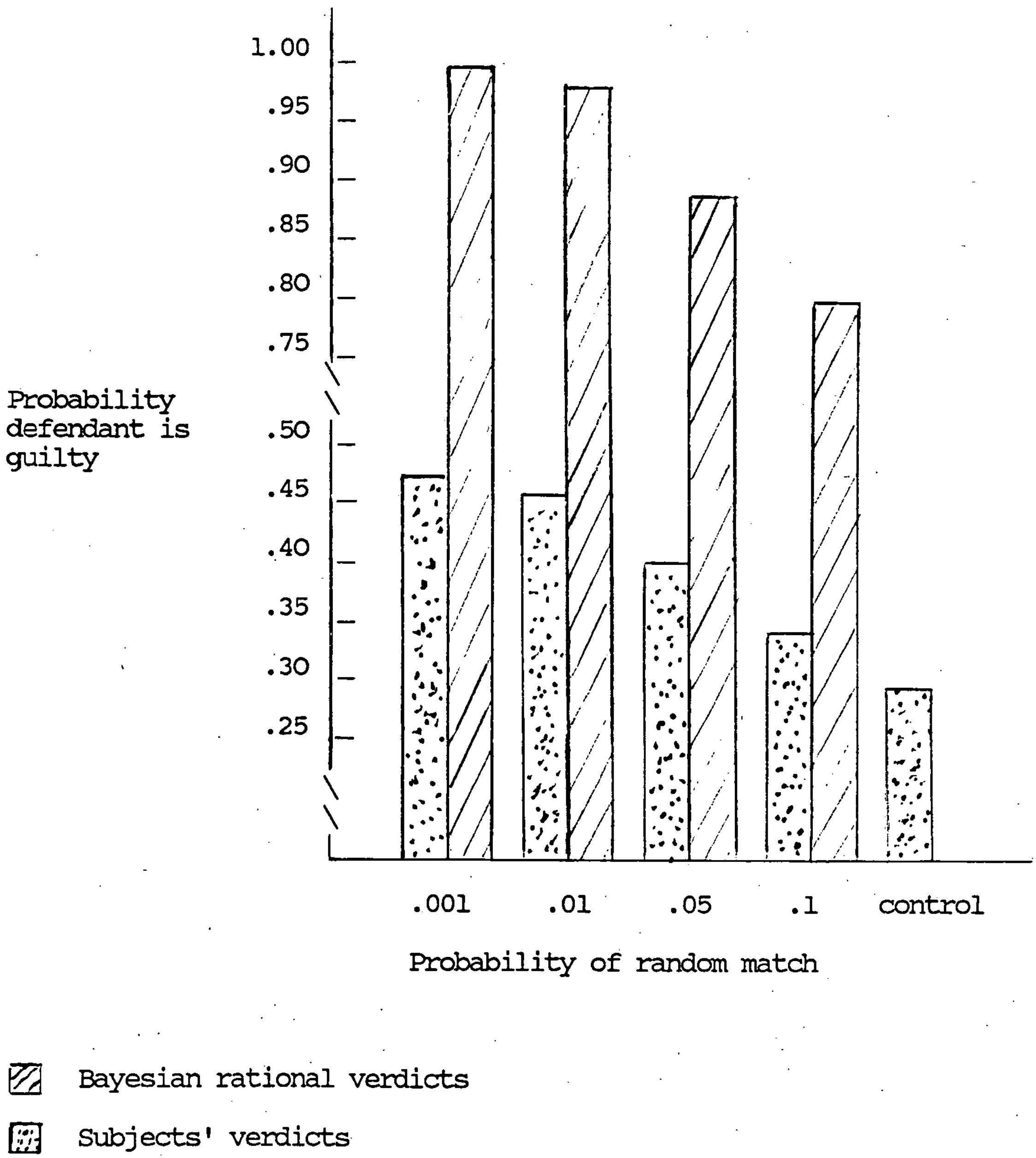
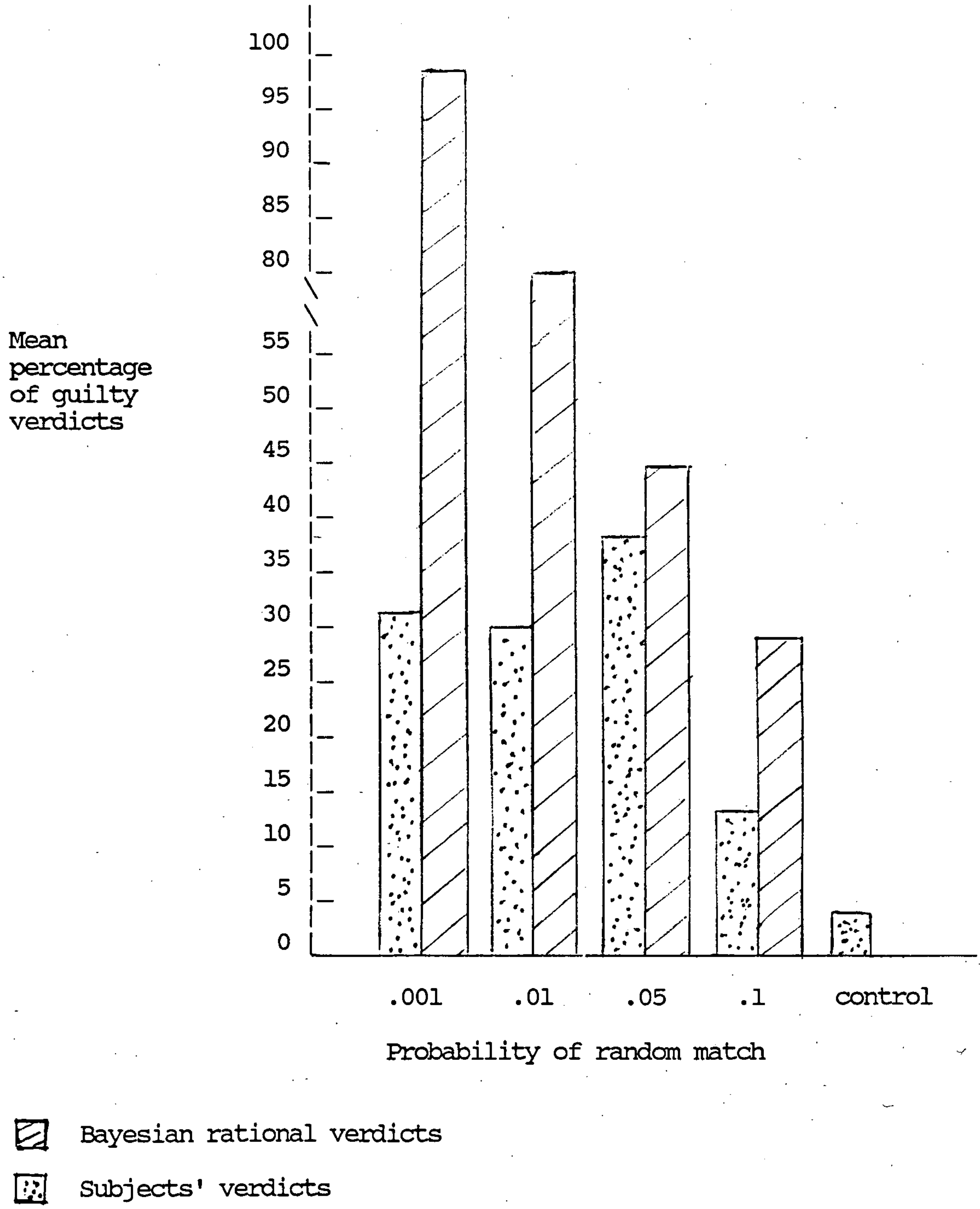


Figure 2. Conviction rate (Percentages) Among Student Mock-Jurors as a Function of Frequency Probability Evidence versus "Rational" Bayesian Conviction Rate.



The percentage of subjects in Groups One, Two, Three, Four and Five who answered this question correctly was 54%, 66%, 80%, 61% and 69%, respectively. These results are presented in Table 10.

Table 10

Percentage of Mock-Jurors who Correctly Answered Quantitative Questions Based on Probabilistic Evidence by Group

	Probability of random match					
Question	.001	.01	.05	.1	control	Grand Mean
No. of people with suspect blood type	54	66	80	62	69	66
Odds of random match	13	23	22	23	53*	27
N =	48	44	45	44	51	232

* "I don't know" was scored as a correct response.

A similar question asked subjects to state the odds that the blood samples gathered at the scene of the crime matched the defendant's blood type. To respond correctly, subjects had to recall the size of the population and the frequency probability evidence, perform a simple division, and rephrase the answer in terms of odds. For example, in one group, subjects were informed that "blood type A2B occurs in 10% of the population." To obtain the correct answer to this question,

subjects simply had to convert the given percentage into an odds statement by dividing the percentage by 100, i.e., 1∅/1∅∅ = 1/10 = one in ten. The correct answer had to be selected from ten forced-choice response alternatives, similar to the previous question. Overall, the error rate for responses to this question was 73%, considerably higher than that for the previous computational question, indicating that subjects have some difficulty in converting simple percentages to odds. Subjects in the group in which the frequency probabilities were lowest produced the most errors. The percentage of subjects in Groups One, Two, Three, Four and Five who selected the correct response was 13%, 23%, 22%, 23% and 53%, respectively. These results are included in Table 10. Note that in the control group, technically, the correct answer is "I don't know", since no frequency probability information was presented which would permit subjects to compute the odds. 53% of the subjects in Group Five selected this response. Excluding the control group, 18% of the subjects in Groups One, Two, Three and Four responded that they did not know the answer to this question (N = 181).

<u>Evidence of the Prosecutor's Fallacy</u>. To determine whether the mock-jurors were susceptible to the "prosecutor's fallacy" (i.e., assumed that the given frequency probability was the probability of defendant's guilt), a count was made of the number of subjects in each group who reported as an estimate of defendant's guilt the complement of the probability of a random match between the defendant's blood-type and that found at the scene of the crime. For example, a count was

taken of the number of mock-jurors in Group Four (10% frequency probability) who assessed defendant as 90% guilty; the number of mock-jurors in Group Three (5% frequency probability) who reported that the defendant was 95% guilty, etc. The number of subjects whose responses could be classified as instances of the prosecutor's fallacy was very low, i.e., this response strategy was not common. Results of this analysis are presented in Table 11.

Table 11

Frequency and Percentage of Subjects by Group who Committed the Prosecutor's Fallacy

Frequency Probability	Percent Committing Prosecutor's Fallacy	N
.1%	2	1
1%	0	0
5%	0	0
10%	4.5	2
Total	1.6	3/182

Weight of the Evidentiary Facts. Two kinds of questions were posed to assess the manner in which subjects perceive and weigh probabilistic evidence in the process of reaching a verdict in a criminal case, and the impact of this evidence on the weight accorded to other evidentiary facts. The one-page summary presenting the case in dispute to subjects

was distilled to 20 core evidentiary facts. Each of these facts was rated by subjects three times. First, the 20 evidentiary facts were rated individually for their incriminating value and their weight. Subjects rated the extent to which each fact indicated that the defendant was culpable by using the negative end of the scale if the fact had no incriminating value, and the positive end of the scale if they considered the fact to be incriminating. The particular value marked on either the positive or negative end of the scale indicated the weight of the item. Grand means and mean responses in each group are reported in Table 11. The facts are presented in rank-order by overall mean according to the incriminating value assigned thereto by subjects. Least incriminating facts, with negative values, are presented first, and the most incriminating facts are found at the end of the list.

One-way analyses of variance were conducted on students' responses to each of the foregoing questions assessing the incriminating weight of the evidentiary facts. For most facts, there were no significant differences in the mean weight assigned to the facts by jurors in each experimental group. This is not surprising since the core facts were invariant in all experimental conditions. However, in response to the question concerning the facts which were varied by experimental group, i.e., regarding the weight of the frequency probability evidence, there was a significant main effect for group, $\underline{F}$ (3, 177) = 4.76, $\underline{p} < .01$. The results of this analysis are presented in Table 13.

Table 12

Mean Weight of Evidentiary Facts by Group

	Experimental Group					
Evidentiary fact	Grand Mean	.1%	1%	5%	10%	control
Old friends	-25	-20	-23	-17	-38	-28
Biking accident	-16	-9	-16	-12	-25	-17
Def. phoned victim	-9	-1	-5	-6	-16	-15
Boise residence	-9	-4	-14	-6	-19	-5
Pop. of 100,000	-2	0	0	4	-7	-5
Sent sec. to check	4	10	1	13	0	-2
Imp. bus. meeting	6	11	6	4	5	6
Victim punctual	7	9	5	8	8	2
Sec. found victim	7	8	11	14	1	0
Def. indicted	10	11	23	12	4	3
Victim not arrive	12	7	15	4	13	9
Victim lived alone	13	10	7	9	30	9
Not victim's blood	15	21	19	19	11	9
Def. blood tested	15	22	15	16	16	5
Business partners	16	22	21	11	11	14
Samples analyzed	19	23	23	22	21	9
Police suspect def.	22	25	26	15	25	18
Def. face scratched	26	38	21	20	23	25
Frequency evidence	34	49	40	30	14	-
Def. blood matched	55	82	47	44	45	-

Post hoc Scheffé procedures showed only Group 1 and Group 4 to be significantly different from each other. All other comparisons were non-significant, $p > .05$.

Table 13

One-Way Analyses of Variance of Weight Assigned by Student Mock-Jurors to Incriminating Frequency Probability Evidence

Source	SS	df	MS	F	p
Groups	30173.48	3	10057.83	4.76	.003
Error	373895.59	177	2112.40		
Total	404069.07	180	2244.83		

The differences in these mean ratings confirmed the experimental hypothesis that subjects would accord the most weight to the probabilistic evidence in the group in which the information was most incriminating (i.e., .1%), and less weight in the group in which the probative value of the frequency probability evidence was weakest (i.e., 10%). Note that because the phrasing of this question had to be modified for mock-jurors in Group Five, who never received any frequency probability evidence, their responses were excluded from the analysis of variance on this question.

Ratings in response to the statement that "Greg's blood type matched the samples of blood taken from Bill's apartment hallway"

revealed that overall, mock-jurors in the groups in which incriminating frequency probability was presented accorded more weight to this evidentiary fact (M = 55%) than to the other 19 facts which were rated. However, one-way analysis of variance yielded no significant differences in group means in ratings of the weight of this fact, F (3, 188) < 5, p > .05. Once again, differences in the phrasing of the question on the match between the defendant's blood type and that of the blood found at the scene of the crime for subjects in Group Five were the basis for excluding Group Five from analyses of variance in responses to this question.

In addition to the series of questions on the relative weight of the facts in evidence, subjects considered the 20 evidentiary facts as a group and selected facts which tended to exonerate or inculpate the defendant. Subjects were asked to divide 100 points among the 20 facts first selecting those indicating the defendant was guilty, and then those which indicated the defendant was not guilty. Not all subjects' scores added to 100. To facilitate analysis, a score of zero was assigned to responses if no weight was accorded to the fact, and a score of one was assigned if some weight, either inculpatory or exculpatory, was accorded to the fact. Unlike the previous question which sought ratings on all twenty facts, here subjects could ignore certain facts, and were instructed to mark, in turn, only those which they believed had either inculpatory or exculpatory value.

Univariate analyses of variance were conducted on all 40 responses. While the likelihood of Type 1 error increases when conducting multiple

tests on dependent variables that are not wholly independent, in this case, subjects had the option of assigning some points to each of the twenty facts which were listed. To mitigate the increased possibility of Type I error, a stringent test of significance was applied, i.e., p of .001 or less.

Of the facts regarded as inculpatory, analyses of variance yielded significant differences between the means of groups in the assessment of only two evidentiary facts: (1) the frequency of the suspect blood type in the Boise population; and (2) the fact that the suspect's blood type matched the samples of blood taken from the victim's apartment hallway. These results are presented in Tables 14 and 15. Post hoc Scheffé analyses for comparing group means revealed that means in Groups One, Two, Three and Four were significantly different from the means in Group Five in both cases, F (4, 228) = 24.14 and 48.54, respectively, $p < .01$. Note that for purposes of these analyses, since responses were scored as either 1 or 0, responses from subjects in the control group were included in the analyses of variance. Specific differences in the wording of the questions for subjects in the control group, in which no frequency information was provided, as opposed to that for subjects in the four experimental groups, in which frequency information was provided, were: (a) "Tests on blood samples found in Bill's apartment hallway were inconclusive" versus "Blood type A2B occurs in 10% of the population of Boise"; (b) "It was impossible to determine whether Greg's blood type matched the samples taken from Bill's apartment hallway" versus "Greg's blood type matched the samples

Table 14

Mean Weight of Incriminating Evidentiary Facts by Group

	Experimental Group					
Evidentiary fact	Mean	.1%	1%	5%	10%	control
Frequency evidence	.77 *	.92	.96	.85	.84	.35
Def. blood matched	.81 *	.96	.93	.95	.96	.31

* p = .001; 1 = inculpatory 0 = not inculpatory

of blood taken from Bill's apartment hallway."

One-way analyses of variance yielded significant differences between group means in student mock-jurors' assessments of five evidentiary facts regarded as indicative of defendant's guilt. The facts which were differentially selected as exculpatory by subjects were: (a) the size of the population in Boise, where the murder took place; (b) the failure of the victim to arrive at work on the day of the meeting; (c) the business relationship of the victim and the suspect; (d) the frequency of the suspect's blood type in the population; (e) the match between the blood samples from the victim's apartment hallway and those gathered from suspect. Grand means and group mean ratings of these facts are presented in Table 16.

Table 15

One-Way Analyses of Variance of Incriminating Weight Assigned by Student Mock-Jurors to Frequency Probability Evidence and Fact That Defendant's Blood Matched Samples from the Scene of the Crime

Frequency Probability Evidence

Source	SS	df	MS	F	p
Groups	12.34	4	3.09	24.14	.001
Error	29.14	228	.13		
Total	41.49	232	.18		

Matching Blood Samples

Source	SS	df	MS	F	p
Groups	16.7	4	4.18	48.54	.001
Error	19.6	228	.09		
Total	36.31	232	.16		

Table 16.

Mean Weight of Exonerating Evidentiary Facts by Group

	Experimental Group					
Evidentiary fact	Mean	.1%	1%	5%	10%	control
Pop. of 100,000	.38*	.35	.57	.46	.38	.17
Victim never arrive	.14*	.06	.16	.2	.27	.04
Business partners	.62*	.77	.66	.56	.73	.41
Frequency evidence	.43*	.18	.43	.51	.49	.59
Def. blood matched	.31*	.1	.18	.17	.16	.86

* $p = < .005$; 1 = exonerating; 0 = not exonerating

Perceived reliability of different types of evidence. Following the questions on the evidentiary facts, using a scale from zero to 100, subjects provided estimates of the reliability of three types of evidence: eyewitness testimony and two types of circumstantial evidence--statistical evidence and blood tests. One-way analyses of variance on each of these dependent variables yielded no significant differences between group means. Mean ratings by group are presented in Table 17 and Figure 3.

The mean reliability rating of eyewitness testimony was 56%; the mean reliability rating of statistical evidence was 60%. By comparison, blood tests were regarded as more reliable than either

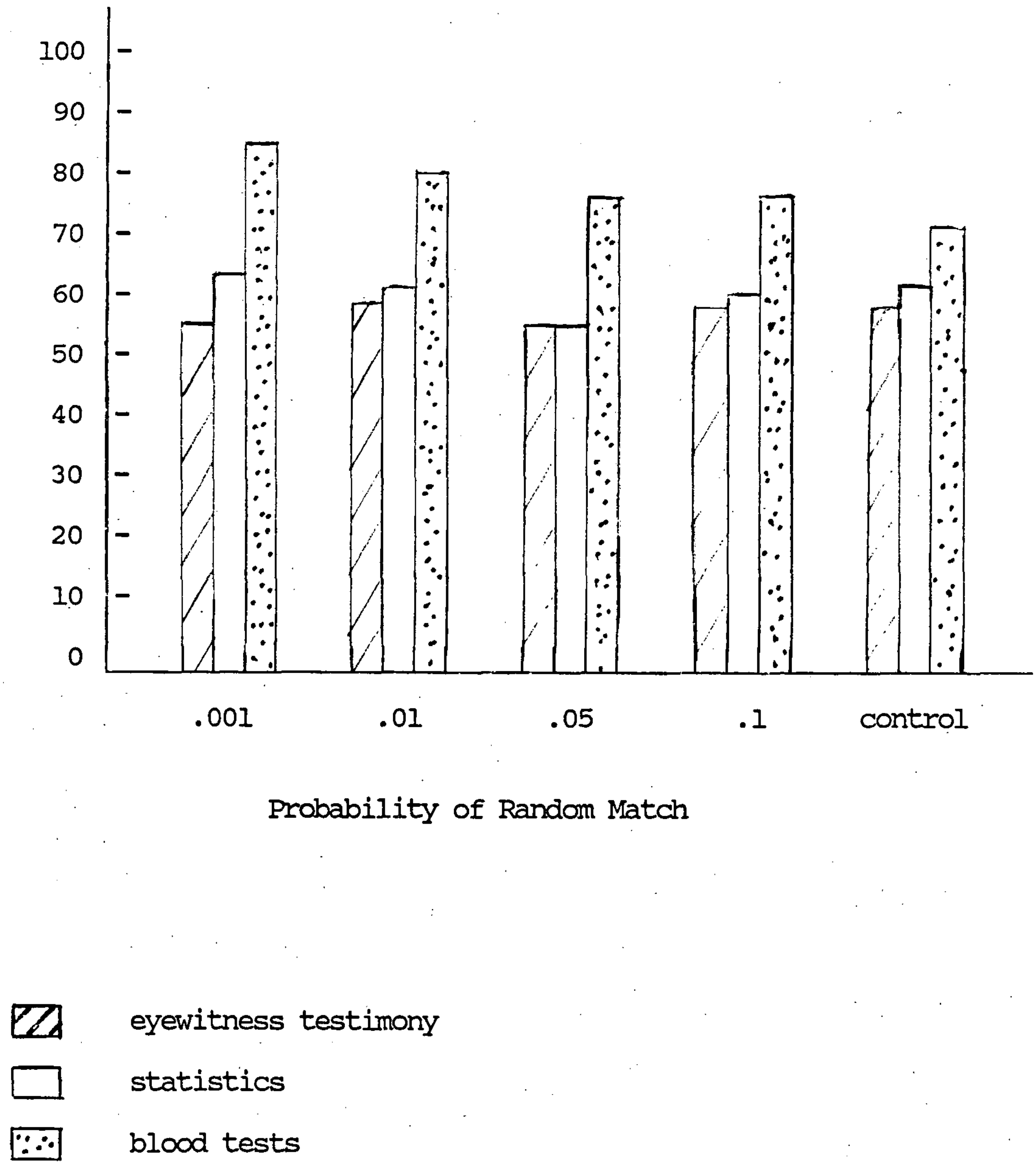
eyewitness testimony or statistical evidence, ($\underline{M}$ = 79%). Mean responses to the question on the reliability of blood tests also varied more in relation to the independent variable. Subjects who received the most incriminating frequency probabilities rated blood tests as most reliable ($\underline{M}$ = 84%). Reliability ratings in other groups diminished steadily as a function of the frequency probability evidence. Mean ratings in Groups Two, Three, Four and Five were 80%, 78%, and 77% respectively.

Table 17

Mean Ratings of Reliability of Eyewitness Testimony, Statistics and Blood-Tests by Group

	Type of Evidence		
Group	Eyewitness Testimony	Statistics	Blood tests
.001	55.5	63.7	84.3
.01	57.0	60.8	80.0
.05	54.6	55.1	78.4
.1	57.0	60.1	77.6
control	57.5	61.2	73.2
Grand Means	56.3	60.2	78.6

Figure 3. Perceived Mean Reliability (Percentage) of Eyewitnesses, Statistics and Blood Tests by Group



Discussion

The Influence of Variations in Frequency Probabilities on Mock-Jurors' Factual Inferences. The student mock-jurors performed surprisingly poorly on the computational questions, given the rudimentary nature of the calculations required to obtain a correct answer. Approximately one third (34%) of the students failed to discern the correct number of people in the suspect pool. One explanation for this finding is that subjects were unable to recall the facts which would have permitted them to deduce the correct answer, i.e., the population of Boise and the frequency of the suspect blood type. In other words, this result may be attributable to poor memory and not poor comprehension. Since a clear grasp of the meaning of the probabilistic evidence is a prerequisite to drawing an appropriate inference regarding the weight to accord the probabilistic testimony, errors at this stage may account for the fact that results in the experimental groups are so undifferentiated.

Considerably more errors were made by students in response to the second deductive task (M = 73%) than in response to the first deductive task (M = 34%). A large number of undergraduate students were unable to convert the frequency probabilities, presented as percentages, into odds statements based on the size of the suspect population. Even when taking into account the fact that only those subjects who correctly deduced the answer to the first question had any chance of answering the second question correctly, the proportion of errors remains high (59%). One interpretation of these results is that where quantitative

evidence is concerned, numerous student mock-jurors are more inclined to guess at the meaning of the evidence rather than to expend the effort necessary to compute the correct answers. Subjects were never instructed to work the answers out in writing. A second possible explanation for these findings is that the subject-jurors did not understand the evidence, or did not know how to calculate the answers to the computational questions. Whatever the precise reasons for the high error rate, it is clear that more detailed, step-by-step questioning of jurors is needed to assess at what stage in the inferential process students' comprehension of probabilistic evidence is weak. Hypotheses about overuse and underuse of scientific evidence cannot be adequately tested if a sizeable number of subjects are reassigning themselves to different experimental groups by virtue of their deductive, computational errors. Both failure to attempt to compute the number of persons in the suspect pool and failure to accurately compute the number of persons in the suspect pool might bias estimates of the weight of the scientific evidence. Had more of the deductions based on the evidentiary facts been accurate, the inferences drawn from the probabilistic evidence might have produced more differentiation in the responses by subjects in the four experimental groups. Of course, weight assigned to the evidence is not a direct measure of inferences drawn from the frequency probability evidence.

The prosecutor's fallacy. The foregoing results provide scant support for the hypothesis that mock-jurors are susceptible to the prosecutor's fallacy. The proportion of subjects whose estimates of

the defendant's guilt coincided with the complement of the frequency probability was negligible, 1.6%. The low incidence of this fallacy may be a result of the fact that few of the subjects interpreted the burden of proof as stringently as 90% likelihood of guilt, whereas all the experimental frequency probabilities ranged between .1% and 10%. However, it is the extreme probability ratios that are generally thought to generate the greatest confusion with the criminal burden of proof. This study provides no support for this hypothesis.

The defense fallacy. Despite the manipulations in the independent variable, which dramatically modified the likelihood of a random match between the defendant's blood and that found at the scene of the crime, differences in the mock-jurors' estimates of defendant's guilt in experimental Groups One, Two and Four, with discrimination probabilities of 1/1000, 1/100 and 1/10 respectively, were slight, and at best, produced a linear rather than an exponential function. For example, notwithstanding a ten-fold increase in the probative value of the incriminating evidence presented in Groups One and Two, estimates of the defendant's guilt in these groups were very close (45% vs. 47%, respectively). Notably, the conviction rates in these groups were within one percent of each other (30% vs. 31%, respectively). These results indicate that the student mock-jurors were insensitive to the variations in the frequency probabilities, or, alternatively, that they found it difficult to accord appropriate weight to the probabilistic evidence.

Some insight into the extent to which the mock-jurors tended to ignore the probabilistic evidence was gained by the comparison of their verdicts with those of the control group and with those of hypothetical Bayesian rational jurors. First, comparisons with the control group revealed that in all four experimental groups, the estimates of guilt and conviction rates were higher than those of mock-jurors who received no frequency probability evidence. This finding indicates that the incidence of the defense fallacy may be overrated. Second, comparisons with the hypothetical Bayesian jurors tended to support the theory that mock-jurors underuse probabilistic evidence in reaching a decision. In all experimental groups, subjects failed to accord sufficient weight to the probabilistic evidence. The magnitude of the disparity between assessments of defendant's guilt by the mock-jurors and those of the Bayesian hypothetical jurors was most pronounced in the groups in which the frequency probabilities were lowest, i.e., most extreme.

Overall, comparisons with Bayesian norms illustrate that subjects are not wholly insensitive to the probabilistic evidence, but that they do not make much distinction between vastly different discrimination probabilities. The findings regarding the weight attributed to the frequency probability evidence show that this value was practically invariant in Groups Two, Three and Four, suggesting that while subjects were aware that this information was probative of the defendant's guilt, and while they clearly regarded this as the most incriminating fact presented, they were insensitive to differences between probabilistic evidence that is fairly incriminating (odds of 1/100) and

probabilistic evidence that is not very probative (odds of 1/10). The reasons for the lack of discrimination are not yet clear. Underuse may be attributable to a number of factors, such as miscalculation of the odds, inability to compute the odds, failure to consider the size of the suspect pool, etc.

Burden of proof and conviction rates. Comparisons of students' verdicts with their numerical estimates of the defendant's guilt indicated that subjects used varying standards in establishing a criterion for returning a guilty verdict, possibly because of variations in their interpretation of the meaning of the instructions on the burden of proof. A substantial number of jurors who returned a guilty verdict believed there was less than a 95% chance that the defendant committed the murder. Of the 23% of the subjects who convicted the defendant, only 5% of the convictions were from subjects who estimated the likelihood of the defendant's guilt to be between 95% and 100%; while 18% of the convictions came from subjects whose estimates of the defendant's guilt were lower than 95%. Jurors are not provided with any numerical criteria to assist them in understanding the burden of proof instructions. Judges typically do not permit the attonreys to provide explanations along these lines to the jurors, for fear of a mistrial. Nonetheless, there are some commentators who advocate this practice. Law-schools typically teach that a "preponderance of the evidence means 51%", and that "beyond a reasonable doubt" means 95-100%. However, these quantitative interpretations of the legal standards are not imparted to jurors.

While many lawyers operate under the assumption that these quantitative values are associated with the burdens of proof, this assumption is not explicitly supported by the case law. Accordingly, jurors who apply a lower burden of proof criterion than 95% cannot technically be declared to have misapplied the law.

While some lawyers and judges may disagree as to whether there is cause for concern should a juror return guilty verdict using an 88% estimate of guilt, and whether this juror might be regarded as unduly conviction prone, few will disagree that the conviction based on an estimate of guilt below 75% conforms to the intended application of the criminal burden of proof. The criterion employed by some students in the present study is clearly lower than that intended by law, resulting in a higher proportion of convictions that would occur if jurors applied the stand which the lawyers have in mind. This phenomenon may be interpreted as a prosecutorial error or bias. To assess whether this bias was independent of the probabilistic evidence, the range of interpretations of the burden of proof by subjects in the control group was compared with that of subjects in the experimental groups. The range of values produced by students in the control group was no different from that of student mock-jurors in the other groups. In all groups, a certain percentage of students returned a guilty verdict using a criterion for the criminal burden of proof, i.e., estimates of guilt between 50% and 90% were considered adequate for a conviction in all five groups.

Presumption of innocence. One interesting finding in relation to the experimental hypotheses was that many facts with no legally incriminating value *per se* receive some inculpatory weight in student's decision-making. In particular, students rated the fact that the defendant was indicted as indicative of guilt. Similarly, the fact that the police officer suspected that the defendant knew more than he would say was perceived as indicative of the defendant's guilt. These results indicate that presumptions of innocence may be difficult to obtain in practice, and that student mock-jurors do not base their decisions on evidentiary facts alone. Information which, based on Constitutional principle and legal presumption, ought to have a neutral value insofar as the culpability of the defendant is concerned was perceived by the subjects as incriminating.

Conclusion

Study One provided some support for the hypothesis that mock-jurors underuse probabilistic evidence in reaching a verdict. No support for the prosecutor's fallacy was found. Similarly, these findings produced no support for the defense fallacy. No support was found for the hypothesis that probabilistic evidence is confused with the burden of proof; however, widespread differences in the interpretation of the legal criterion complicate analyses of this matter.

The results partially replicated those reported by Faigman (1983) in that only in the group in which the most extreme frequency probabilities were presented did college students accord the evidence

more appropriate weight. These results extend the findings of Faigman by showing that at the extreme end of the spectrum, just as at the non-probative end of the scale, students fail to make the necessary distinctions in accordance with the changing probabilities.

CHAPTER FOUR

VARYING THE FORM OF FREQUENCY PROBABILITIES IN CIVIL AND CRIMINAL CASES

Study Two

Purpose and Rationale

The foregoing review and the findings in Study One led to the conclusion that mock-jurors are susceptible to inferential errors in reaching a verdict in cases in which probabilistic information is included in the evidence. First, student mock-jurors tended to ignore or underuse probabilistic information in comparison with Bayesian norms, and failed to distinguish between markedly different incriminating probabilities. Second, subjects misused the probabilistic information. Many subjects did not appear to understand fully the significance of the evidence in concrete terms, i.e., error rates were high in response to questions about the number of individuals possessing the suspect characteristic within a given population, and about the odds of a random match. The generally poor performance by the students raised questions about the competency of potential jurors, who often have less education or less exposure to mathematics and statistics than do college students. Adult registered voters called for jury duty may be less competent in drawing appropriate inferences from probabilistic information, and more susceptible to the prosecutor's fallacy or the defense fallacy than are college students.

The subjects' apparent inability or unwillingness to perform simple computations to convert the probability statement to an expression of the odds of a random match raises a question about the influence of the form of the presentation of the probabilistic evidence. In Study One,

frequency probability information was presented in percentages by specifying the incidence rate of the suspect blood type. The implication of the information in terms of the odds of a random match (between the defendant's blood type and that of the blood found at the scene of the crime), given the size of the population, was never specified. The presentation format in which the implications of the probabilistic information are not explicitly stated may have deterred some subject-jurors from inferring the significance of the probabilistic evidence and consequently, from assigning the appropriate weight to the evidence. If subjects are informed of the likelihood of a random match, as opposed to having to infer this information, they may be more likely to use this information appropriately in evaluating the defendant's culpability.

As noted in Chapter Two, numerous techniques to enhance the comprehensibility of probabilistic evidence have been recommended, but have not been empirically validated. Included among the recommendations were suggestions to simplify the language by means of which the expert witness conveys the probabilistic information to the mock-jurors. The impact of variations in probabilistic language used by the expert witnesses remains unclear. Prior research has shown that subjects were more fascile with probabilities expressed in percentages and odds as opposed to fractiles and log odds, but that overconfidence persisted even when odds and percentages were used (Lichtenstein, Slovic & Phillips, 1982). In addition to suggestions to enhance juror comprehension of probabilistic information by modifying linguistic variables, a frequent recommendation is to include illustrations to

enliven statistical presentations. The impact of visual aids in the courtroom to illustrate or emphasize probabilistic evidence is untested.

Aside from questions about the influence of the form of the probabilistic evidence upon mock-jurors' decisions, the issue of possible confusion between the burden of proof and probabilistic evidence was not satisfactorily resolved by Thompson's (1984, 1985) studies. One way to test whether mock-jurors confound the burden of proof with the probabilistic evidence is to present the same substantive testimony to mock-jurors while varying the burden of proof or decision-rule. Characterization of a case as either a civil or a criminal matter has more to do with the desired outcome of the lawsuit than with the facts upon which it is premised. For instance, a fatal automobile accident may come to trial as either a criminal matter in which the driver is prosecuted for vehicular homicide, or as a civil action for wrongful death. Minor procedural features vary depending upon whether the case is civil or criminal. For example, the moving party in a criminal case is typically the state or the federal government; in a civil case, typically a private plaintiff. The final decision to be made by jurors in criminal as opposed to civil cases also varies. In criminal cases, jurors must determine the defendant's guilt, in civil cases, the defendant's liability.

The Thompson (1984, 1985) studies indicated that while more mock-jurors were likely to ignore the scientific evidence, some mock-jurors overused this information. Thompson (1984) attributed this finding to confusion between the probabilistic evidence and the burden of proof.

Another plausible explanation for these findings is that jurors make differential use of the evidence based on their own attitudes towards and experiences with quantitative information. In other words, by examination exclusively of task variables, important information about biases which jurors bring to the task, which may also exert an influence over their decision-making, may be ignored. A study which explores attitudes and individual differences of the mock-jurors may be instrumental in accounting for the somewhat disparate findings reported in previous research on this topic.

In particular, mock-jurors' educational and occupational backgrounds may influence their decisions. When mathematical evidence is presented, a juror well-acquainted with statistical and mathematical concepts is likely to respond differently from a juror having little or no mathematics experience. A juror with mathematics anxiety may resort to different strategies from those employed by non-anxious mock-jurors in evaluating mathematical evidence. Mathematics anxiety may also impede mock-jurors' understanding of and ability to use mathematical information appropriately. Mock-jurors with different levels of expertise in mathematics may respond differently to variations in the presentation of the evidence.

With these ideas in mind, the purpose of Study Two was to explore further the nature of mock-jurors' inferences when presented with probabilistic information in the context of either a civil or criminal trial. Three specific goals were: (a) to determine whether mock-jurors who receive probabilistic qualitative information tend to confuse this numerical evidence with the quantitative burden of proof, and whether

variations in the burden of proof affect the conviction rate; (b) to determine whether mock-jurors make fewer inferential errors when probabilistic information is presented in the form of percentages than when it is presented in the form of odds, and to assess the impact of this variation upon perceived culpability of the defendant; (c) to determine whether mock-jurors make fewer inferential errors when scientific information is accompanied by a bar-graph illustration than when no illustration is provided, and to assess the impact of this variation upon perceived culpability of the defendant.

Hypotheses

A. Civil versus Criminal Burden of Proof:

The major hypothesis concerning the influence of the burden of proof was that mock-jurors in criminal and civil conditions would render similar estimates of defendant's guilt, but that mock-jurors in the former group would return fewer verdicts finding the defendant guilty of the alleged crime than would mock-jurors in the civil condition. This hypothesis was based on the legal presumption that mock-jurors required to apply the more stringent burden of proof will be more likely to find that the evidence did not rise to the criminal threshold (commonly paraphrased as 90% likelihood of guilt) than will mock-jurors whose task is to apply the less stringent civil threshold (commonly paraphrased as 51%). An alternative hypothesis, which did not presume that mock-jurors would uniformly interpret the burden of proof as intended by the law, was that there would be no significant variation in the number of arson verdicts as a consequence of

variations of the burden of proof.

Finally, it was hypothesized that mock-jurors rendering a verdict in a criminal case would be more susceptible to the prosecutor's error than would mock-jurors in a civil case.

B. Odds versus percentages

Precise hypotheses about differences in mock-jurors' errors and estimates of guilt as a consequence of variations in the linguistic form of the probabilistic evidence were more difficult to formulate in the absence of much prior research on this issue. The general prediction was that percentages would facilitate comprehension of the statistical evidence, and that mock-jurors would find the odds formulation more difficult. If one adopted the judicial perspective that mock-jurors overuse probabilistic evidence, mock-jurors whose comprehension is increased should accord less weight to the evidence than mock-jurors whose comprehension is not increased. On this basis, one would predict that estimates of defendant's guilt in the percentage group would be lower than estimates of defendant's guilt in the odds group. However, there is little empirical support for this position, aside from studies by Thompson (1984, 1985) of the prosecutor's fallacy, and even these do not reveal a tendency to overuse the evidence by a majority of the subjects. Thompson found more subjects who tended to ignore the probabilistic evidence. Faigman's (1983) findings also indicate more undersuse of probabilistic evidence. Thus, prior research favors the hypothesis that mock-jurors tend to ignore statistical evidence. Accordingly, increases in the comprehensibility of the probabilistic evidence should result in increased use.

Therefore, a major hypothesis concerning the influence of this variable was that mock-jurors assigned to the percentage condition would return higher estimates of culpability than mock-jurors assigned to the odds condition.

C. Bar graph versus no bar graph

The hypothesis concerning the influence of providing some mock-jurors with a bar graph illustrating the results of scientific tests was similar to that of presenting the probabilistic information in percentages; i.e., it was predicted that the estimates of culpability by subjects who received the illustration would be higher than those by subjects who received no illustration. This prediction was based on the following premises: First, the chart should serve to emphasize the scientific testimony, causing subjects to pay more attention to the probabilistic evidence than subjects who consider the evidence in the absence of any illustration. Both the additional attention and/or the fact that the chart may render the testimony more comprehensible should result in more weight being accorded to this evidence than will be accorded in the absence of an illustration. Second, the chart might enhance mock-jurors' understanding of the scientific testimony, making them less likely to ignore it.

Method

Subjects

Subjects were potential and experienced jurors on call at the King County Superior Court. Registered voters resident in King County are eligible for jury duty, and are called for a two-week term of service. Their names are selected randomly from the voter-registration list. Jurors may request to be excused from duty on certain grounds. Those who are not excused gather in the jury room on the ninth floor of the courthouse every weekday, and wait to be called to trial. Approximately 200 jurors are in attendance daily, however, the administrators increase the pool of jurors when lengthy or controversial trials are scheduled. On any given day, some jurors on duty serve on - civil or criminal trials, while others wait in the jury room to be called on a random basis as needed. Not all jurors who are called from the jury room are actually selected to serve on a trial. To serve on a case, jurors must survive voir dire and peremptory challenges or challenges for cause (bias) by the attorneys presenting the case. If a trial on which a juror is selected to serve lasts fewer than ten days, the juror returns to the jury room and waits out the rest of the term of duty, during which he or she may be called out a second time. Consequently, subjects in this experiment consisted of both inexperienced potential jurors (69%) and experienced jurors (31%) on call in the King County Courthouse jury room. Participant jurors, numbering 223 in all, included slightly more females than males (109 versus 102), and 13 subjects who did not disclose their gender. Approximately one half of the subjects were over 45 years old, born

before 1941. Most subject-jurors had completed high school (93%), with about 80% continuing their schooling after high school. Approximately 14% attended vocational schools, 20% completed a two-year college degree, 30% completed a four-year college degree, and 16% earned either a masters' degree or doctorate. 60% had studied mathematics since attending high school, while 37% had taken some statistics courses. Subjects received no payment for their participation in the study.

Materials

A written trial summary was prepared with the assistance of forensic scientists on the staff at Western Washington State Crime Laboratory in Seattle, Washington. The National Institute of Trial Advocacy civil arson case, Flinders v. Mismo, used in mock trials to train lawyers in trial techniques (Beskind, Bocchino, Ordover & Seckinger, 1983) comprised the basic source materials. These materials contain controverted facts, some of which support a finding that a fire which destroys an industrial plant is accidental, while other facts support a finding that the fire was the result of arson. The factual information is carefully balanced so that a verdict for either party is plausible.

For this study, two versions of the case materials were prepared. In one version, the facts were presented in the context of a civil lawsuit, in which the owner of the destroyed plant sues his insurance carrier to recover the proceeds of the insurance policy, and the major defense to liability is that the fire was deliberately set. In the second version, the facts were presented in the context of a criminal trial, in which the state prosecutes the owner of the plant for

suspected arson.

The experimental materials consisted of a 4,000-word written summary of the following five trial components: (a) lawyers' opening statements; (b) the case in-chief for arson; (c) the defense against allegations of arson; (d) lawyers' closing arguments; and (e) instructions from the judge to the jury.

The testimony was presented in narrative form, summarizing points made during direct and cross-examination of each of six trial witnesses. Three witnesses testified in favor of arson: the company's former bookkeeper, the Vice-President of the bank which refused to finance the company's ventures; and the Chief Fire Marshal. Three witnesses were called in support of the case for an accidental fire: the owner of the company, his administrative secretary, and a fire investigator.

Three trial exhibits referred to in the witness examinations were appended to the materials: (a) a plan of the first floor of the industrial plant; (b) an excerpt of the insurance policy showing the dates and amounts of increases in coverage to a total of $1,667,000; (c) a summary of the results of chromatographic tests performed by the expert witness on four gasoline samples. A copy of the experimental materials showing variations in the independent variables, and the follow-up questionnaire, is attached as Appendix B.

Design

A 2x2x2 between-subjects factorial design was employed, in which the first variable was burden of proof. The two levels of this variable were (a) civil version, in which the burden of proof is a

"preponderance of the evidence"; and (b) criminal version, in which the burden of proof is "beyond a reasonable doubt." The second independent variable was the linguistic form of the probabilistic testimony presented by the expert witness. This variable also had two levels: (a) likelihood of a random match expressed in terms of odds; and (b) likelihood of a random match expressed in terms of a percentage. The third factor that was varied was the presence of a graph summarizing the results of the tests which comprised the basis of the probabilistic scientific testimony: (a) a bar graph illustration was included on one of the exhibits; (b) the bar graph illustration was omitted.

The trial materials in each experimental condition were identical in all respects with the following exceptions:

(a) Civil versus criminal versions: There were two differences between the civil and criminal trial materials: identification and posture of the parties to the litigation and the burden of proof. In the civil version, the company owner was identified as the plaintiff, and the insurance company was identified the defendant. In the criminal version, the State of Washington was identified as the prosecutor, and the company owner was identified as the defendant. The burden of proof instruction which accompanied the civil version was a preponderance of the evidence, while the burden of proof instruction which accompanied the criminal version was beyond a reasonable doubt. Mock-jurors receiving the civil version were required to determine whether the insurance company was liable. Mock-jurors receiving the criminal version were required to determine whether the company owner was guilty. The purpose of this manipulation was to study the effect

of varying the burden of proof, a task variable, rather than the effect of varying the consequences of the decision, a context variable. To maintain minimal realism, and to avoid questions from mock-jurors familiar with the legal system who know that a civil burden of proof does not apply in a criminal case and vice-versa, minimal lexical changes to the materials were made where necessary, e.g., "guilty" versus "liability", "prosecutor" versus "plaintiff". Aside from these necessary lexical modifications, an effort was made to control the potential influence of other contextual information known to influence mock-jurors' decisions. The case provided no basis to infer that the defendant was a repeat-offender or a social menace. There was no discussion of penalty or sentence recommendation in the criminal condition. There was no requirement that mock-jurors decide how much money to award the prevailing party in the civil condition.

To control for order of presentation effects, the order in which the information was presented to subject-jurors in any of the experimental conditions was invariant. Note that while in an ordinary civil trial, the plaintiff's case is generally presented before that of the defense, in certain instances, such as when an affirmative defense is alleged, the defendant bears the burden of production, or the burden of going forward with that evidence. Accordingly, in an insurance case in which the affirmative defense of arson is raised, to avoid having to rebut unknown evidence, the plaintiff typically elects to wait until the defense offers its proof of arson before the plaintiff presents facts in rebuttal. In other words, the party bearing the burden of production presents its case-in-chief first. This arrangement was

particularly suited to the goals of this experiment, as it minimized the chances of introducing a possible confounding variable into the experiment by varying the order of presentation of witnesses, which would be the result if the typical plaintiff-defendant presentation sequence were followed. Thus, in the civil version, the plaintiff presents his rebuttal case to allegations of arson following the proof offered by the insurance company in support of the case for arson. One subject who happened to be married to an attorney inquired about the reason for the departure from the typical order of presentation.

(b) Odds versus percentages: The second independent variable, the probabilistic testimony, was introduced in this trial by the Chief Fire Marshal who described the results of chromatographic tests performed on matching gasoline samples taken from the site of the fire and from the gas tank of the suspected hired arsonist. Half of the subjects received a version in which the crucial probability statement by the expert in asserting that the gases had a common origin was: "The odds against this result occurring by chance alone are 1 in 1000." The other half of the subjects received trial summaries in which the probability statement by the expert read: "The likelihood of obtaining this result by chance alone is .1%" In each version, corresponding language appeared on Exhibit B, a half-page verbal summary of the results of the chromatographic tests. Subjects in the odds condition read: "The odds against samples 2 and 4 matching by chance alone: 1 in 1000." Subjects in the percentage condition read: "Likelihood that samples 2 and 4 match by chance alone: .1%."

(c) Bar graph illustration versus no graph: The third independent variable, a graphic illustration to support the probabilistic testimony, was introduced by adding to half of the experimental materials a bar graph on the lower half of the page marked "Exhibit B", appended to the trial summaries. The illustration consisted of a black-and-white reproduction of a slide used as an exhibit in a 1984 arson trial by a forensic expert when testifying that gas chromatography tests formed the basis for his opinion that the gasoline residue at the fire scene matched automotive gas found in the trunk of the suspect's car. (See Appendix C, attached.) The information contained in the graph itself did not vary across experimental conditions. The graph showed four gasoline samples compared on each of twelve attributes, and did not directly illustrate the frequency probability statement made by the expert witness.

Dependent Measures

A written questionnaire, seven pages in length, was designed to elicit four types of responses from subjects: (a) forced-choice responses, in which two or more alternative answers were presented; (b) supply questions in which subjects were asked to provide brief, one-word answers by filling in a blank; (c) rating questions which sought a number on a scale from 1 to 100, indicating either credibility, likelihood of occurrence, confidence, etc. (d) handwritten responses of one or two sentences.

A number of verdict measures were used: Subjects provided a verbal (written) verdict (guilty/not guilty and liable/not liable); and an estimate of the defendant's culpability on a scale of zero (not guilty)

to 100 (definitely guilty). In addition, subjects rated their confidence in their verdicts, and provided information about the facts perceived as most important, and major reasons for their decision.

A few comprehension and recall questions were included to ascertain how well subjects grasped the essential probabilistic evidentiary facts presented in the trial summary. For example, they were asked to state the value of the insurance policy in issue, the date on which loan payments were due, how much money the company had borrowed, the number of gasoline samples that were tested, etc. Subjects were also asked to provide examples of direct and circumstantial evidence and to state which instruction from the judge they regarded as the "most important instruction", without further qualification.

To gather information about the weight accorded to the evidence, on a scale from zero to 100, subjects provided ratings of the credibility of each witness, the reliability of scientific evidence, chromatography tests, eyewitness testimony, and expert testimony. Measures of the relative weight accorded to each of the witnesses were obtained by asking subjects to divide 100 points among the six witnesses.

To determine how subjects used the contested facts provided in a fragmented presentation via witness testimony in direct and cross-examination, to form a coherent script or picture of what happened, they were asked to provide a rating on a scale from zero to 100, indicating the likelihood of occurrence of 27 events. Nine items favored an arson interpretation, nine favored an accident interpretation, nine were neutral on this issue. The items included uncontroverted evidentiary facts, controverted evidentiary facts,

inferred facts and ultimate facts to be decided by the mock-jurors.

Some illustrative examples of statements which favored an interpretation of arson are listed below:

(a) An uncontroverted evidentiary fact: Likelihood that chance factors account for the similarity of gasoline samples taken from the scene of the fire and from the gas tank of the car belonging to the purported arsonist.

(b) A controverted evidentiary fact: Likelihood that a witness overheard a discussion in which the purported arsonist was described as "a torch."

(c) An inferred fact: Likelihood that the similarity of the gasoline samples from the scene of the fire and from the gas tank of the purported arsonist indicated arson.

(d) An ultimate fact: Likelihood that the fire was caused by arson.

Finally, subjects provided some demographic information and then completed a 47-item math attitude questionnaire drawn from the Fennema-Sherman Math Attitude Scales, (Fennema, 1976). The scales, consist of six positively-stated and six-negatively stated items with five response alternatives: strongly agree, agree, neutral, disagree, strongly disagree. The split-half reliability for each scale was $\geq .89$ (Fennema & Sherman, 1979). Items from four scales were randomly distributed into one instrument, viz., the math-confidence scale (12 items), the math-anxiety scale (12 items), the math-usefulness scale

(11 items) and the effectance-motivation scale (12 items). (The latter is designed to provide some indication of subjects' persistence in solving math problems.)

Procedures

Following roll-call in the jury room at the King County Courthouse, potential jurors who were not assigned to a particular courtroom were invited by the jury administrator to participate in a University of Washington study. Those who indicated interest in the project were told it would require about one hour of their time. They were handed consent forms (a sample is attached marked Appendix D) and randomly assigned to one of eight experimental conditions, for which they were provided stimulus materials and an attached response questionnaire. They were informed that their participation in the study would not prevent them from serving on a real trial if they were called out of the room before they completed the questionnaire. (This happened several times). Mock-jurors were instructed to work on their own and not to discuss the materials with others. Monitors collected the completed questionnaires. No time limit was prescribed. Mock-jurors took between 1 and 2 hours to complete the task. Participants were assigned to experimental conditions as reported in Table 18.

Table 18

Study Two: Number of Subjects Assigned to Experimental Conditions

Version	Experimental group				Total
	Odds		Percentage		
	Graph	No graph	Graph	No graph	
Civil	28	24	28	31	111
Criminal	27	26	30	29	112
Total	55	50	58	60	223

Results and Discussion

Reported results are based on analyses of responses from 223 subjects. One protocol was discarded because of extensive missing data.

Verdict

Three-way analyses of variance on mock-jurors' verdict responses produced an interaction between the three independent variables, showing that a finding of arson was more likely in the civil condition in which the probabilistic evidence was expressed in odds, and not accompanied by any bar-graph illustration, than in a criminal case in which the expert expressed the probabilities as percentages and used an

illustration in support of the circumstantial evidence. F (1, 213) = 4.42, $p < .05$. See Table 19. Note that a value of 1 represents a verdict to the effect that the fire was the result of arson planned by the owner of the aluminum company, i.e., a verdict of guilty in the criminal version, and a verdict of no liability in the civil version. A value of 2 represents a verdict that arson was not proved, i.e., a verdict of not guilty in the criminal version, and a verdict of liability in the civil version. The cell means showing the interaction between the independent variables are presented in Table 20.

Table 19

Three-way Analysis of Variance of Mock-Jurors' Ultimate Verdicts Regarding Arson

Source	SS	df	MS	F	p
Burden of Proof	.13	1	.13	.54	NS
Bar Graph	.05	1	.05	.22	NS
Linguistic Form	.17 .35	1	.17	.67	NS
BOP x Bar Graph	.00	1	.00	.01	NS
BOP x Ling.	.73 92	1	.73	2.98	NS
Bar Graph x Ling.	.19	1	.19	.78	NS
BOP x Ling. x BG	1.54	1	1.54	6.26	.01
Residual	52.23	213	.25		
	2.83	7	.25		
Total	55.06	220	.25		

25
220) 5506
440
1166
7)283

Table 20

Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof Regarding Mock-Jurors' Ultimate Verdicts of Arson

		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	1.57	1.29	1.43
	Criminal	1.54	1.63	1.59
		1.56	1.46	

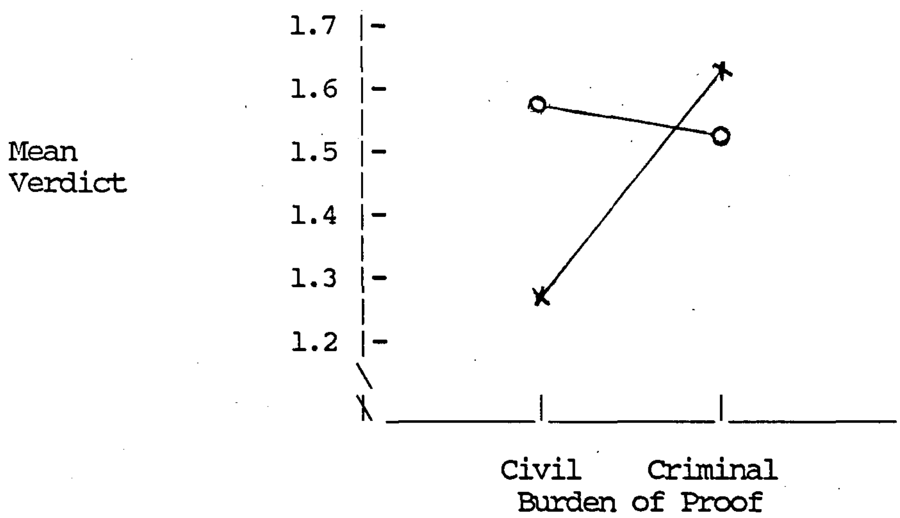
Linguistic Form of Probabilistic Evidence: Odds

		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	1.50	1.68	1.59
	Criminal	1.59	1.47	1.53
		1.55	1.58	

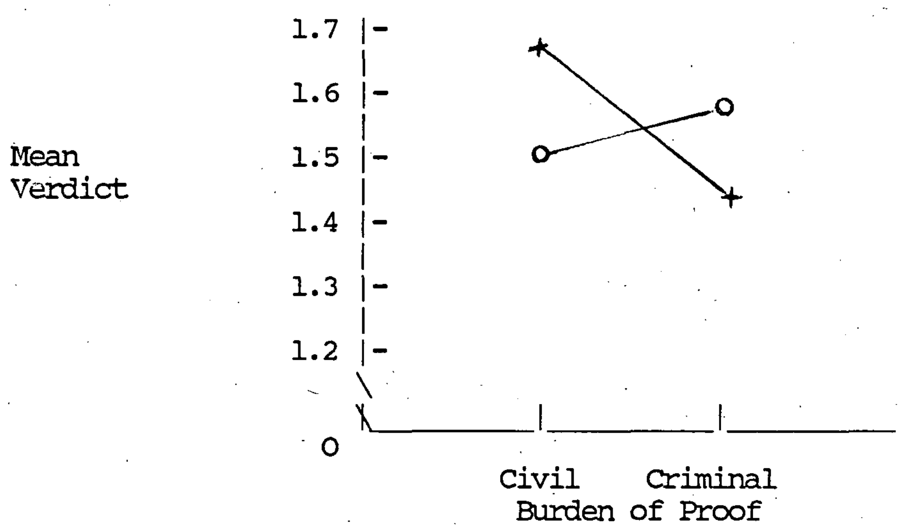
Linguistic Form of Probabilistic Evidence: Percentages

Key: 1 = arson verdict 2 = verdict that arson was not proved

Figure 4. Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof Regarding Mock-Jurors' Ultimate Verdicts of Arson



Linguistic Form of Probabilistic Evidence: Odds



Linguistic Form of Probabilistic Evidence: Percentages

Key: o--o Visual aid x--x No visual aid

1 = arson verdict 2 = arson not proved

Analyses of variance were also conducted on the mock-jurors' probability estimates that fire in the warehouse was the result of arson, and probability estimates that the fire was accidentally caused. There were no significant differences attributable to the independent variables regarding estimates that the fire was deliberately caused. However, estimates that the fire was accidental yielded a significant two-way interaction: $\underline{F}$ (1, 198) = 4.92, $\underline{p}$ = 0.03. Results of this analysis are presented in Table 21. Group means in the civil versus criminal condition were 42% and 38% respectively. Mean estimates that the fire was accidental in the illustration versus no illustration condition were 41% and 38% respectively; and in the odds versus percentage condition, were 37% and 43% respectively.

Table 21

Three-way Analysis of Variance of Mock-Jurors' Estimates of the Likelihood the Fire was Accidentally Caused

Source	SS	df	MS	F	p
Burden of Proof	645.31	1	645.31	.66	NS
Bar Graph	238.53	1	238.53	.24	NS
Linguistic Form	1916.61	1	1916.61	1.97	NS
BOP x Bar Graph	231.36	1	231.36	.30	NS
BOP x Ling.	234.88	1	234.88	.24	NS
Bar Graph x Ling.	449.78	1	449.78	.46	NS
BOP x Ling. x BG	4798.23	1	4798.23	4.92	.03
Residual	193022.52	198	974.86		
Total	201620.41	205	938.51		

Table 22

Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof in Mock-Jurors' Estimates that the Fire was Accidental

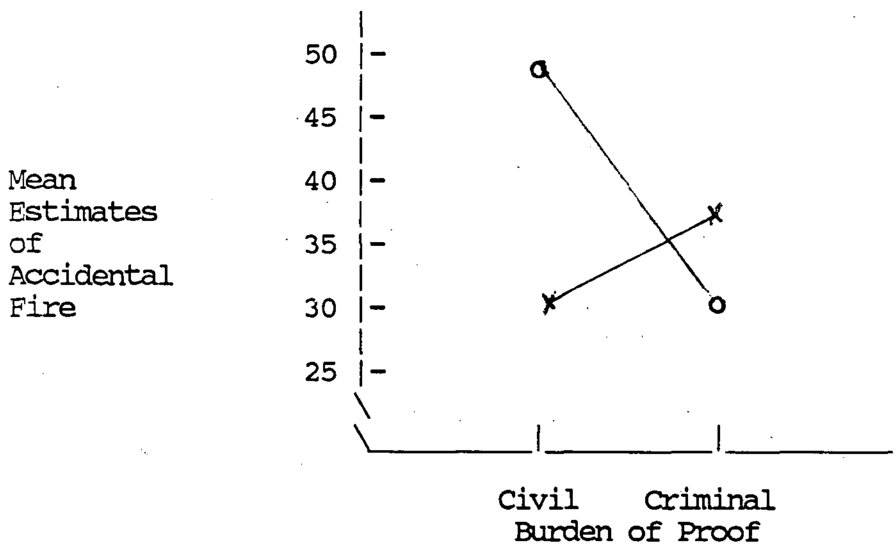
		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	49.32	31.54	40.43
	Criminal	30.80	37.22	34.01
		40.06	34.38	

Linguistic Form of Probabilistic Evidence: Odds

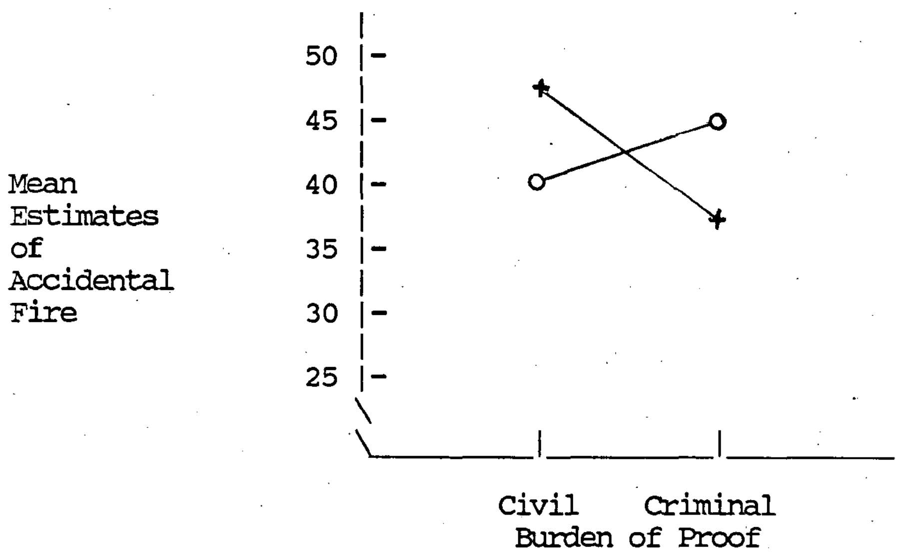
		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	40.18	47.92	44.05
	Criminal	45.56	38.8	42.18
		42.87	43.36	

Linguistic Form of Probabilistic Evidence: Percentages

Figure 5. Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof On Mock-Jurors' Estimates the Fire was Accidental



Linguistic Form of Probabilistic Evidence: Odds



Linguistic Form of Probabilistic Evidence: Percentages

Key: o--o Visual aid x--x No visual aid

Mock-jurors' inferences about the extent to which the scientific evidence increased the likelihood that Jackson was guilty varied considerably, depending upon the experimental condition to which subjects were assigned. Mock-jurors were asked to provide an estimate of the likelihood that the similarity between the gasoline samples taken from the storeroom floor and from Avery's gas tank indicated arson. Three-way analysis of variance revealed a significant three-way interaction, as shown in Table 23. $\underline{F}$ (1, 198) = 12.67, $\underline{p}$ < .01. The scientific evidence was rated evenly by subjects in the civil ($\underline{M}$ = 52.4%) and criminal ($\underline{M}$ = 52.6%) conditions, but as more incriminating when the probabilities were expressed as odds ($\underline{M}$ = 54.6%) as opposed to percentages (50.6%). Mock-jurors' who were provided with a visual aid showing the similarity of the four gases on the chromatographic tests were less prosecutorial ($\underline{M}$ = 54.8%) than subjects who received no bar graph illustration (50.1%). The cell means are displayed in Table 24.

Verdict and Weight of the Evidence

Overall, the mock-jurors were divided fairly evenly in determining whether the fire was the result of arson. Ignoring experimental condition, 47% of the verdicts returned reflected a decision that the fire was caused by arson. On the average, mock-jurors were 68% confident that their verdicts were correct. In addition to the dichotomous verdict measures, mock-jurors' numerical estimates that the fire was attributable either to arson or accidental causes were analyzed. The mean estimate of the likelihood of arson was 59%, while the mean estimate of the likelihood of an accidental fire was 40%.

Table 23

Three-way Analysis of Variance of Mock-Jurors' Estimates of the Likelihood that Matching Gasoline Samples Imply Arson

Source	SS	df	MS	F	p
Burden of Proof	1.48	1	1.48	.00	NS
Bar Graph	1292.31	1	1292.31	1.52	NS
Linguistic Form	969.43	1	969.43	1.14	NS
BOP x Bar Graph	306.37	1	305.37	.36	NS
BOP x Ling.	1345.21	1	1345.21	1.58	NS
Bar Graph x Ling.	356.17	1	356.17	.42	NS
BOP x Ling. x BG	10754.43	1	10754.23	12.67	.001
Residual	168131.94	198	849.15		
Total	183336.12	205	894.32		

To assess whether mock-jurors tend to confuse the burden of proof with the frequency probability evidence (in this case a .1% likelihood of a random match between the gasoline samples) responses to two questions were compared: (a) An estimate the extent to which the matching gasoline samples indicated that the fire was caused by arson (i.e., this question sought an inference based only on the probabilistic evidence); and (b) a more general question which asked mock-jurors to state the likelihood that the fire was deliberately caused. The latter question was a paraphrase of the penultimate factual issue before the mock-jurors. In other words, mock-jurors' final verdicts would differ from responses to this question if they

Table 24

Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof in Mock-Jurors' Estimates of the Likelihood that Matching Gasoline Samples Imply Arson

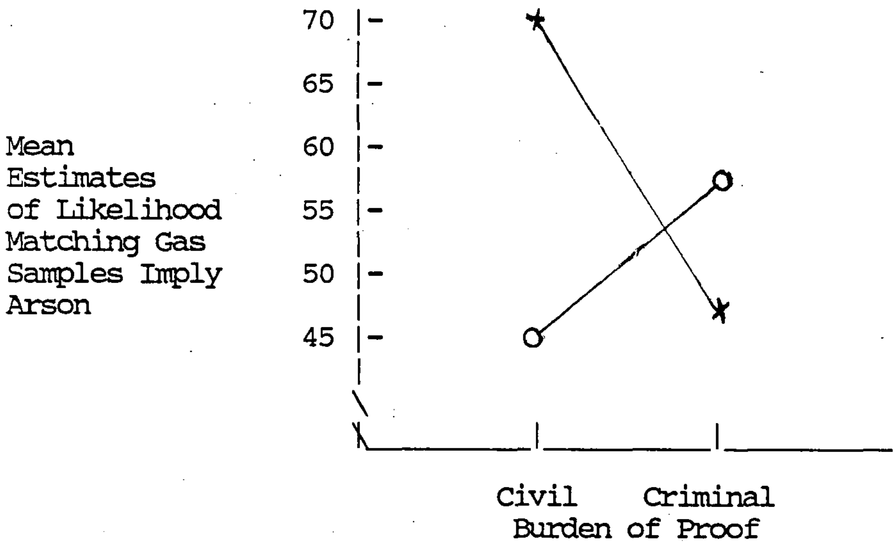
		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	44.09	69.81	56.95
	Criminal	57.00	47.96	52.48
		50.55	58.89	

Linguistic Form of Probabilistic Evidence: Odds

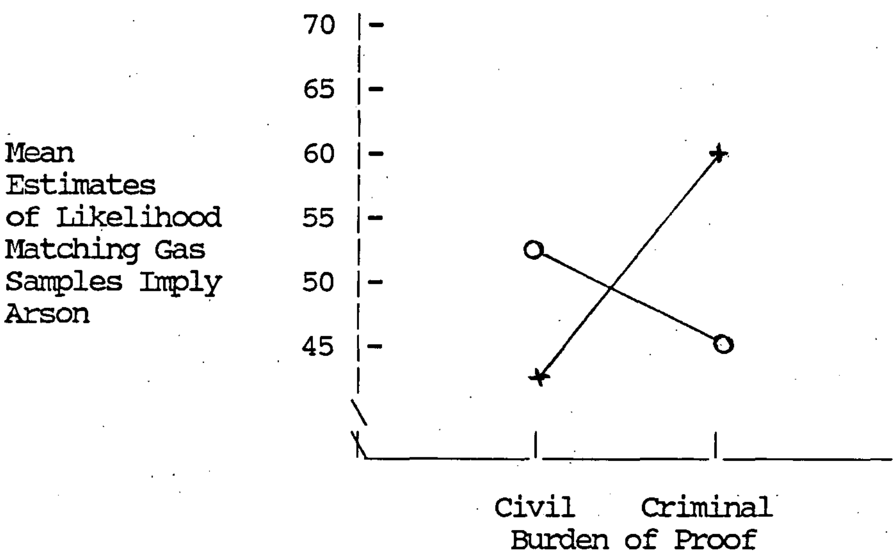
		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	52.32	43.42	47.87
	Criminal	46.11	60.40	53.26
		49.22	51.91	

Linguistic Form of Probabilistic Evidence: Percentages

Figure 6. Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof in Mock-Jurors' Estimates of Likelihood that Matching Gasoline Samples Imply Arson



Linguistic Form of Probabilistic Evidence: Odds



Linguistic Form of Probabilistic Evidence: Percentages

Key: o--o Visual aid x--x No visual aid

found that the fire was the result of arson, but did not find that Arthur Jackson knowingly and maliciously instigated the fire. In responding to the question whether the fire was deliberately caused, mock-jurors were not limited to a consideration of the incriminating value of the probabilistic evidence, and were free to consider any other facts in evidence which might support a finding of arson, such as the financial trouble in which the company found itself, the unusual timing of the dramatic increase in insurance coverage, the unusual circumstances surrounding the hire of Avery, etc. Estimates of guilt were regarded as instances of the prosecutor's fallacy if responses of 99% or more were returned, although, strictly speaking, only responses of 99.99% qualify for this designation. Overall, 8% of the mock-jurors produced estimates of defendant's culpability as high as 99% to the first question, and 15% of the mock-jurors gave estimates as high as 99% or more in response to the second question. Only mock-jurors in the first group can be regarded as potentially susceptible to the prosecutor's fallacy.

Perceived Reliability of the Evidence

The independent variables influenced the extent to which subjects perceived gasoline frequency studies to be reliable. Three-way analysis of variance on ratings of the reliability of gasoline frequency studies yielded a two-way interaction among all three independent variables, a significant interaction between the linguistic form of the probabilistic evidence and the presence or absence of a visual aid, and a main effect for burden of proof. The analysis of variance results and cell means are displayed in Tables 25 and 26.

Group means were higher in conditions in which no illustration was provided ($\underline{M}$ = 66% vs. 65%) and in which the probabilistic testimony was expressed as a percentage ($\underline{M}$ = 66% vs. 65%). $\underline{F}$ (1, 184) = 4.55, $\underline{p}$ < .05. See Figures 7 and 8. Perceived reliability of gasoline frequency evidence was significantly higher among mock-jurors assigned to the civil case condition than among those assigned to the criminal case condition; $\underline{M}$ = 70% and 61%, respectively. ($\underline{t}$ (193) = 2.6, $\underline{p}$ < .01.)

Table 25

Three-way Analysis of Variance of Mock-Jurors' Ratings of the Reliability of Gasoline Frequency Evidence

Source	SS	df	MS	F	p
Burden of Proof	4013.20	1	4013.20	6.68	.01
Bar Graph	28.45	1	28.45	.05	NS
Linguistic Form	67.33	1	67.33	.11	NS
BOP x Bar Graph	513.60	1	513.60	.86	NS
BOP x Ling.	1372.18	1	1372.78	2.29	NS
Bar Graph x Ling.	3573.53	1	3573.53	5.95	.02
BOP x Ling. x BG	2731.26	1	2731.26	4.55	.03
Residual	110433.66	184	600.18		
Total	122599.25	191	641.88		

Table 26

Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof in Mock-Jurors' Estimates of the Reliability of Gasoline Frequency Studies

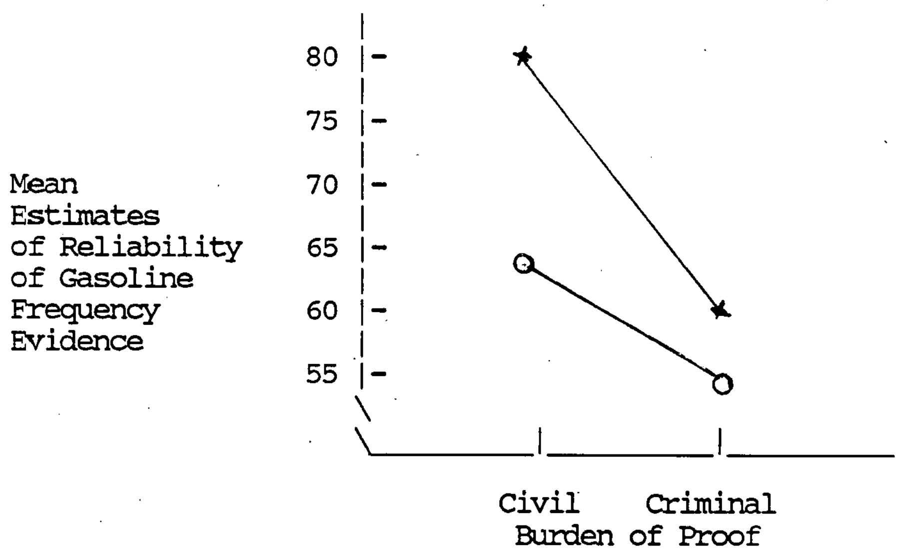
		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	64.21	79.55	71.88
	Criminal	55.00	60.58	57.79
		59.60	70.07	

Linguistic Form of Probabilistic Evidence: Odds

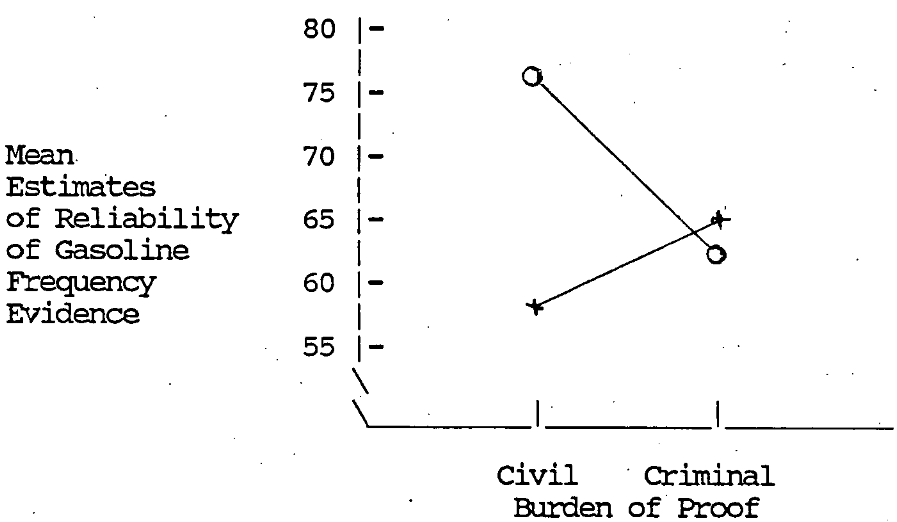
		Visual aid: Bar graph	Visual aid: No graph	
Burden of proof	Civil	76.21	58.18	74.05
	Criminal	62.56	65.20	63.88
		69.39	61.69	

Linguistic Form of Probabilistic Evidence: Percentages

Figure 7. Two-way Interaction Among Linguistic Form of the Probabilistic Evidence, Presence or Absence of Visual Aid and Burden of Proof in Mock-Jurors' Estimates of Reliability of Gasoline Frequency Evidence



Linguistic Form of Probabilistic Evidence: Odds



Linguistic Form of Probabilistic Evidence: Percentages

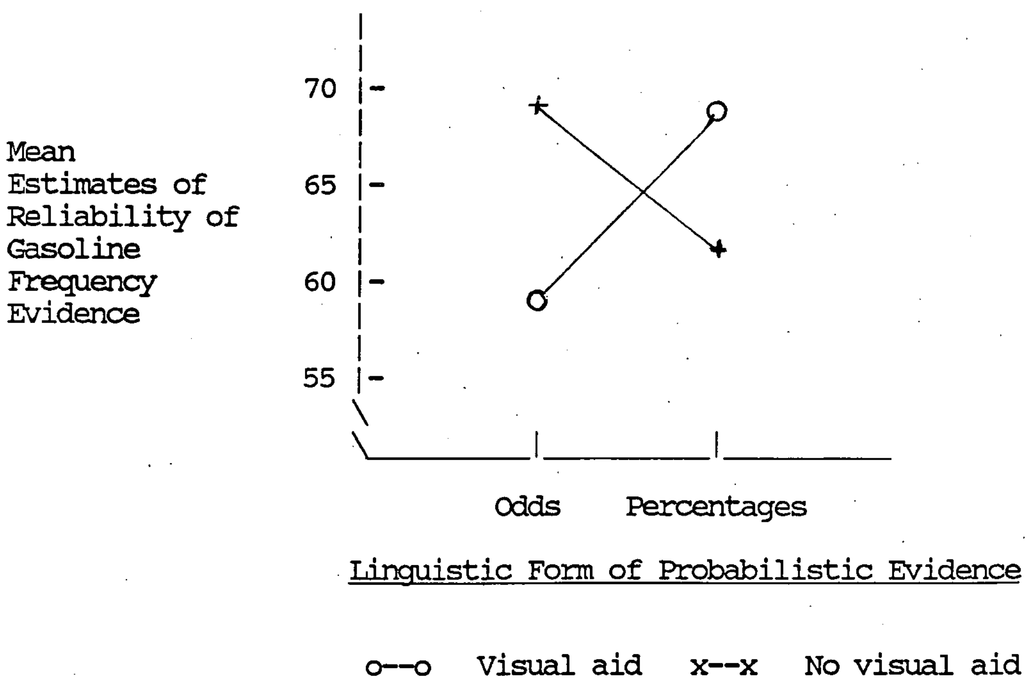
o--o Visual aid x--x No visual aid

Table 27

One-Way Interaction Between Linguistic Form of the Probabilistic Evidence and Presence or Absence of Visual Aid in Mock-Jurors' Estimates of the Reliability of Gasoline Frequency Studies

		Linguistic form of probabilistic evidence: Odds	Percentages	
Visual aid	Bar graph	59.27	69.63	64.45
	No graph	69.27	61.91	65.59
		64.27	65.77	65.44

Figure 8. One-Way Interaction Between Linguistic Form of the Probabilistic Evidence and Presence or Absence of Visual Aid in Mock-Jurors' Estimates of the Reliability of Gasoline Frequency Studies



To determine whether the scientific evidence was accorded undue weight, mock-jurors rated the reliability of various types of evidence. Overall, mock-jurors regarded expert testimony as significantly more reliable than laywitness testimony ($\underline{M}$ = 73% vs. 54%), $\underline{t}$ ($\underline{df}$ = 198) 13.45, $\underline{p}$ = < .01. Direct evidence was considered significantly more reliable than circumstantial evidence ($\underline{M}$ = 78% v. 50%), $\underline{t}$ ($\underline{df}$ = 201) 15.07, $\underline{p}$ = < .01. Chromatography tests were rated as significantly more reliable than gasoline frequency studies ($\underline{M}$ = 73% v. 66%), $\underline{t}$ ($\underline{df}$ = 187) 4.64, $\underline{p}$ = < .01.

Credibility and Weight of Witness Testimony

Three-way analyses of variance were conducted on weight scores assigned to all six witnesses whose testimony was considered by the mock-jurors. Because the likelihood of Type I error increases when multiple tests are conducted on dependent variables that are not wholly independent, and subjects divided 100 points among 6 witnesses, a stringent test of significance was applied, i.e., $\underline{p}$ of .001 or less. No significant differences were found.

Mock-jurors rated the credibility of each of the six witnesses whose testimony they considered. Three-way analyses of variance yielded no significant differences in the group means attributable to the independent variables. Finally, mock-jurors provided estimates of the reliability of witness testimony. Three-way analysis of variance of ratings of the reliability of lay witness testimony revealed a significant interaction between two independent variables, linguistic form of probabilistic information and presence or absence of a visual aid. $\underline{F}$ (1, 184) = 6.56, $\underline{p}$ < .01. These results are presented in

Tables 28 and 29. Laywitness testimony was rated most favorably in conditions in which no illustration was provided (M = 55% vs. 53%) and in which the probabilities were expressed as odds (M = 56% vs. 52%).

Table 28

Three-way Analysis of Variance of Mock-Jurors' Estimates of the Reliability of Lay Witnesses

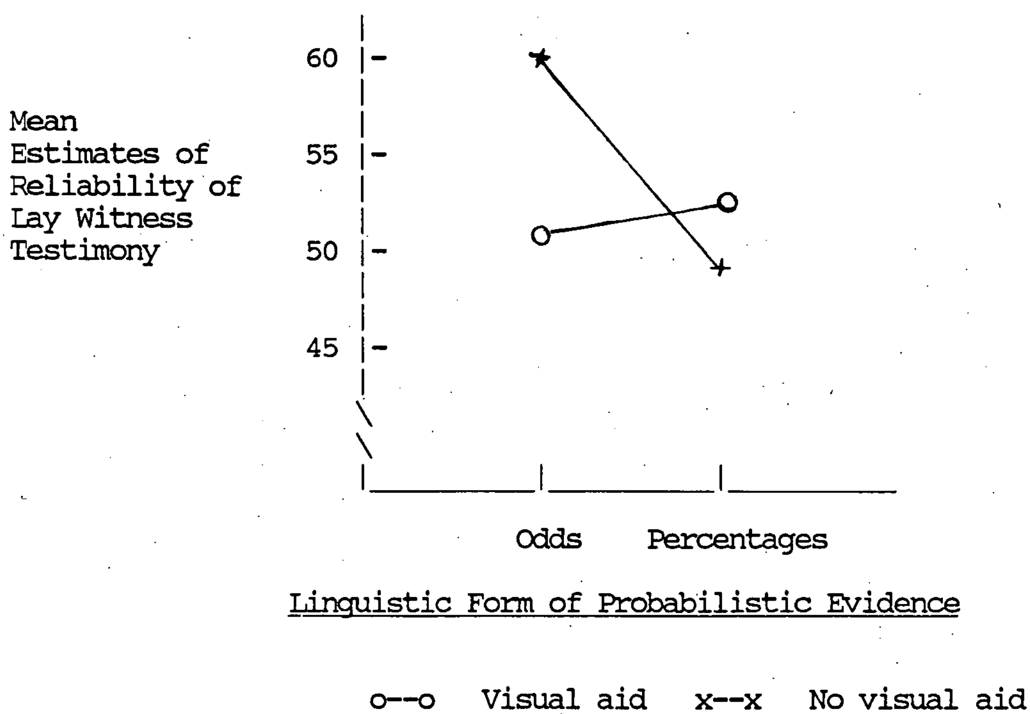
Source	SS	df	MS	F	p
Burden of Proof	136.54	1	136.54	.50	NS
Bar Graph	130.26	1	130.26	.47	NS
Linguistic Form	841.07	1	841.07	3.05	NS
BOP x Bar Graph	293.28	1	293.28	1.06	NS
BOP x Ling.	.00	1	.00	.00	NS
Bar Graph x Ling.	1810.45	1	1810.45	6.56	.01
BOP x Ling. x BG	581.53	1	581.53	2.11	NS
Residual	50800.97	184	276.09		
Total	54647.92	191	286.12		

Table 29

One-Way Interaction Between Linguistic Form of Probabilistic Evidence and Presence or Absence of Visual Aid in Mock-Jurors' Estimates of the Reliability of Lay Witness Testimony

		Linguistic form of probabilistic evidence: Odds	Percentages	
Visual aid	Bar graph	51.59	53.39	52.49
	No graph	59.79	49.47	54.63
		55.69	51.43	53.65

Figure 9. One-Way Interaction Between Linguistic Form of Probabilistic Evidence and Presence or Absence of Visual Aid in Mock-Jurors' Estimates of the Reliability of Lay Witness Testimony



Perceived Credibility of the Witnesses

Confirming evidence for the high credibility ratings assigned to the expert witnesses was evident in the mean credibility scores assigned by mock-jurors to the six witnesses in the case at hand, whose testimony was considered in reaching a verdict regarding arson. The two expert witnesses, Olson and Pinkus, were regarded as more credible than other laywitnesses, with the exception of the Vice-President of the bank from whom Jackson sought refinancing. Results are presented in Table 30.

Table 30

Mean Credibility Ratings (Percentages) of Witnesses by Mock-Jurors

Witness in support of arson	Mean Credibility
Former bookkeeper, Marie Williams	45
Vice-President of bank, Robert Anderson	75
Fire Chief (expert), Howard Olson	72
Witness in support of accidental fire	**Mean Credibility**
Company owner, Arthur Jackson	45
Administrative assistant, Sonia Peterson	43
Fire investigator (expert), David Pinkus	65

0 = Extremely unbelievable   100 = Extremely believable

Two-tailed t-tests revealed that Olson (M = 73%), the expert who presented the scientific evidence and probabilistic testimony, was rated as significantly more credible than Pinkus (M = 66%), t (df = 202) = 3.47, $p < .01$. Pinkus presented no quantitative information.

Relative Weight of the Testimony

In addition to credibility ratings of each witness scale, mock-jurors provided information about the relative weight assigned to each witness by dividing 100 points among the six witnesses whose testimony they considered. Responses from 24 subjects added up to less than 100 or more than 100. For these subjects, the weights were first converted to percentages and then included in the overall analysis. The mean relative weight rankings for the six witnesses are presented in Table 31.

Influence of the Burden of Proof

Handwritten responses summarizing the reasons for final verdicts were provided by 192 mock-jurors. Of this group, 30% (57/192) of the responses made specific reference to the burden of proof or to the sufficiency of the evidence in meeting the decision-criterion. Thirty-four subjects (17%) specifically incorporated words such as "preponderance" or "beyond a reasonable doubt" in their answers. Subjects in the criminal condition (M = 20%) were more likely to make specific mention of the language than were subjects in the civil condition (8%).

Table 31

Mean Relative Weight Assigned to Witness Testimony by Mock-Jurors

Witness in support of arson	0:100
Former bookkeeper, Marie Williams	16
Vice-President of bank, Robert Anderson	17
Fire Chief (expert), Howard Olson	23

Witnesses in support of accidental fire	0:100
Company owner, Arthur Jackson	16
Administrative assistant, Sonia Peterson	12
Fire investigator (expert), David Pinkus	20

Illustrative examples of responses mentioning the criminal burden of proof are provided in Table 32; examples of responses mentioning the civil burden of proof are provided in Table 33.

Table 32

Sample Mock-Juror Verdict Justifications Mentioning Beyond a Reasonable Doubt

There is circumstantial evidence to point to arson, but insufficient to prove case beyond reasonable doubt.

I guess I had too many unanswered questions to say that he had Avery beyond a reasonable doubt.

I do not feel the state proved without a reasonable doubt that Jackson hired an arsonist.

An additional 10% of the subjects were clearly engaging in a quantitative as opposed to a qualitative evaluation of the evidence, despite the absence of words such as "preponderance of the evidence" or "burden of proof". The quantitative focus was apparent in phrases such as "not enough evidence", or other remarks pertaining to the sufficiency of the evidence. Sample responses in which the sufficiency of the evidence is mentioned without a specific reference to the applicable burden of proof are listed in Table 34.

Table 33

Sample Mock-Juror Verdict Justifications Mentioning Preponderance of Evidence

The physical evidence presented by the defendant established a preponderance of evidence in favor of the defendant.

There was a preponderance of evidence that the owner was desperate, had tried to cover bases (i.e., insurance increases), had arranged a means to perpetrate the act, and the gasoline types and evidence were compatible to such an act.

The circumstantial evidence was quite strong--enough to convince me based on the preponderance of the evidence test.

Responses reflecting consideration of the burden of proof occurred primarily when subjects did not return a verdict in favor of arson. Seventy-nine per cent (15/19) of the responses from subjects in the civil condition who mentioned the burden of proof were from subjects who did not find adequate evidence of arson. Similarly, 82% of the responses from subjects assigned to the criminal condition who mentioned the burden of proof were from subjects who did not find

Table 34

Sample Mock-Juror Verdict Justifications Evaluating Sufficiency of the Evidence Without Specific Mention of the Burden of Proof

> Although the majority of the evidence points to a guilty verdict, there were too many factors contributing to doubt.
>
> Insufficient evidence presented by the prosecution.
>
> The defendant's case presented facts and testimony that was too persuasive to be mere coincidence.
>
> There was enough circumstantial evidence against Mr. Jackson for a conviction.
>
> Defendant failed to meet its burden.
>
> From the evidence presented it seems more likely that the fire started accidentally.
>
> Although I have my suspicions, I don't feel that adequate evidence has been presented to say that Mr. Jackson torched his plant.

adequate evidence of arson. This parallel pattern of findings emerged despite the fact that the burdens of proof differ, and despite the fact that the final verdict in the civil version for arson was negatively expressed, i.e., the insurance company was held "not liable", whereas in the criminal version, a finding of arson was affirmatively expressed, i.e., Jackson was found "guilty".

Not all mock-jurors' written justifications made reference to the burden of proof or the sufficiency of the evidence. Some 70% of the responses focused more on qualitative aspects of the evidence, providing insight into the scripts or stories reconstructed by mock-jurors of the events in issue. Examples of justifications with a more

qualitative emphasis are listed in Table 35.

One of the purposes of this study was to examine the influence of variations in the burden of proof upon the susceptibility of mock-jurors to the prosecutor's fallacy in cases in which probabilistic evidence is introduced. The experimental hypothesis was that mock-jurors in the criminal case condition would be more susceptible to the prosecutor's fallacy than mock-jurors assigned to the civil condition, in which the burden of proof is lower, thus more readily distinguishable from the numerical values presented in the

Table 35

<u>Sample Mock-Juror Verdict Justifications with Qualitative Evidentiary Discussion</u>

Gut feeling after reading both sides.
The girlfriend was jealous and vindictive, enough so as to accuse her boss of arson, particularly after she was fired. Jackson was in a bad financial position but had hopes of new designs that would again establish his company as sound.
Arthur Jackson's 13-year effort to build a business not likely to be destroyed with the prospect of new business near.
Do not believe an arsonist would be killed. The gasoline samples were not same. Understand other jealous woman.
No link between Avery and act of arson. Williams' testimony inclined to be biased. Williams' testimony re overhearing word "torch" not relevant. Expert witness re arson was very plausible - credible. Not understand why after five hours of conflagration the can of gasoline in the storage room was available as sample.
I believe Jackson's debt, his hiring of Avery and the gasoline spills in the fire point to arson. I question Avery's credentials and the "project" he was working on.

probabilistic evidence. Only 8% of the subjects were potentially susceptible to this error, and there was no indication that assigment to either the civil or the criminal case condition increased the likelihood that mock-jurors committed this error. The qualitative responses from mock-jurors in the form of written justifications for their final verdicts support this finding for only 6% of the mock-jurors appeared to accord the probabilistic evidence undue weight. Accordingly, no support was found for the hypothesis that incidence of this fallacy would be higher among subjects assigned to the criminal condition.

These results raise doubt about the validity of concerns that mock-jurors are susceptible to the prosecutor's fallacy in cases in which probabilistic evidence is presented. It is possible that Thompson (1984, 1985) found a higher incidence rate of this fallacy because his experimental procedures did not permit him to distinguish between inferences based upon the scientific evidence and ultimate factual inferences, i.e., final estimates of culpability.

While variations in the burden of proof had no significant effect upon verdict, this independent variable influenced the way in which mock-jurors viewed the incriminating scientific evidence. When a higher threshold of proof was imposed, mock-jurors were more skeptical about the reliability of gasoline frequency tests, and also accorded significantly more weight to the expert who presented no scientific evidence himself, but who cast doubt on the viability of the evidence proffered by the expert who presented the probabilistic scientific testimony. These "spillover" results provide some, albeit indirect

verification that subjects are not wholly insensitive to variations in the burden of proof.

The second experimental hypothesis was that mock-jurors applying the more stringent criminal burden of proof would return fewer findings of arson than mock-jurors applying the civil burden of proof to the same set of facts. There was no confirming evidence to support this hypothesis. In reaching a final verdict, mock-jurors did not distinguish between the applicable legal thresholds of proof. However, a considerable number of mock-jurors referred to the burden of proof in justifying their verdicts (either by specifically mentioning the applicable standard, or by engaging in a quantitative evaluation of the evidence) indicating that mock-jurors did not simply ignore this judicial instruction. In fact, 37% of the mock-jurors regarded the burden of proof as the most important jury instruction they received, and no other instruction received such frequent mention. Evidence that mock-jurors did apply some kind of test to the factual findings they made was apparent in the disparity between the number of verdicts in favor of arson (47%) and the number of estimates that the fire was caused by arson which exceeded 50%, i.e., 60% of the mock-jurors believed there was more than a 50% likelihood that the fire was the result of arson. Thus, even though, on the average, most mock-jurors found that the evidence supported a finding of arson (the mean estimate was 59% likelihood of arson) some mock-jurors in this group did not find the evidence adequate to meet the requisite burden of proof.

One possible explanation for the null finding is that mock-jurors' interpretations of the legal standards varied considerably within

groups. In other words, some mock-jurors assigned to the criminal condition may have interpreted the burden of proof too stringently (Type I error), while others may have interpreted the burden too loosely (Type II error). Similarly, some mock-jurors interpreting the preponderance instruction may have approximated the criminal standard of proof (Type I error), while others may have loosely applied the civil standard (Type II error). A mixture of Type I and Type II errors within each experimental condition may have produced the virtually indistinguishable group means. Further investigation is needed to provide a full explanation for this finding. These results replicate those recently reported by Sutton-Barbere, Teitelbaum and Johnson (1986) based on a study in which burden of proof was similarly varied in the context of a mock-trial based on a wrongful death arising from a pedestrian-automobile accident.

The Influence of Variations in the Form of Probabilistic Evidence

Three-way analysis of variance determined that mock-jurors in the odds condition were significantly less confident in the accuracy of their verdicts (M = 66%) than were mock-jurors in the percentage condition (M = 71%). $F(1, 213) = 4.31$, $p < .05$. When the contribution to the variance by mathematics anxiety was removed, subjects in the percentage condition were shown to be even more confident in their verdicts (M = 73%), while the confidence-level of subjects in the odds condition did not change (M = 66%). $F(1, 166) = 7.03$, $p < .01$.

The experimental hypothesis concerning variation in the linguistic form of the probabilistic evidence was that comprehension would be

facilitated by the more common form of quantitative usage, percentages, and that as a consequence, mock-jurors who typically underuse probabilistic evidence would pay more attention to it, increasing the estimates of guilt in the percentage group. No significant differences in verdict and estimates of culpability were found as a consequence of this manipulation, thus this hypothesis was not confirmed. Nonetheless, subjects in the percentage condition were significantly more confident of their verdicts, possibly because this format was more familiar.

The Influence of Presence of Absence of a Visual Aid.

To assess the prominence of the bar graph illustration, mock-jurors were asked which trial exhibit was the most important. Three documents were entered into evidence: the insurance contract, the floor-plan of the aluminum factory, and the summary of the scientific tests on the gasoline samples. Several other tangible items were also mentioned in the testimony, without being entered into evidence. Mock-jurors did not always distinguish between items in evidence and items referred to by witnesses which were not in evidence. Approximately 16% of the subjects in both the bar graph illustration condition and the no bar graph condition responded to this question by mentioning items not admitted into evidence, such as Avery's body, the gasoline can found at the scene of the fire, etc. Irrespective of experimental condition, overall, 55% per cent of the subjects mentioned Exhibit B, which contained a summary of the scientific evidence, as the most important exhibit; 11% mentioned the floor plan of the aluminum factory, and 11%

mentioned the excerpt from the insurance policy. Subjects who received the bar graph illustration were more likely ($\underline{M}$ = 60%) to select Exhibit B as the most important exhibit than were subjects who did not receive the bar graph illustration of the chromatographic tests ($\underline{M}$ = 50%). When no bar graph illustration was presented, more subjects selected the insurance policy as more important (5% versus 16%). When no illustration was presented, subjects were also more likely to respond that none of the exhibits was important (9% versus 5%).

The experimental hypothesis concerning the use of the bar-graph to illustrate the results of the comparison of the gasoline samples, which provided a basis for the probabilistic testimony was, not surprisingly, that this visual aid would facilitate comprehension. The presence of a bar graph did not accomplish this, however.

When the variations in the presentational format of the probabilistic evidence were crossed with the civil versus criminal trial versions, mock-jurors' mean estimates of arson based on the scientific evidence did not conform to the experimental hypothesis. However, some indirect verification of the hypothesized effects of variations in the form of the evidence emerged regarding dependent variables other than estimates of the likelihood of arson. For example, there was an increase in the weight accorded to the expert who testified that accidental causes of the fire could not be dismissed when the probabilistic testimony which he opposed was expressed as a percentage and accompanied by a bar graph illustration, and when this testimony was presented in the context of a criminal as opposed to a civil trial. Spillover effects of this nature underscore the

importance of using a variety of dependent measures in addition to perceived culpability or ultimate verdict.

The finding that subjects who received the illustration tended to accord less weight to the expert who presented this testimony is another example of a spillover effect. Perhaps the illustration served to demystify the scientific evidence. This interpretation is consistent with the finding that mock-jurors in the no illustration group viewed the scientific circumstantial evidence as more indicative of arson. The multiple regression analyses indicate that of the three independent variables, the presence or absence of the illustration had the most influence, though extremely small, on the extent to which juror's viewed the probabilistic evidence as incriminating.

Because so many subjects were confused about the meaning of the terms "circumstantial" evidence and "direct" evidence, the relationship between the use of an illustration and subjects' reliability ratings of these forms of evidence cannot readily be interpreted.

Comprehension and recall of facts:

Several questions tested mock-jurors' understanding and recall of the evidence presented in the trial summary. Errors were fairly common. Approximately 30% of the mock-jurors answered each of the factual questions (about the amount of the insurance coverage, the date when the insurance was usually increased, how soon company loans were due, etc.) incorrectly.

First, mock-jurors were asked about the extent of the aluminum company's financial trouble. From the stated facts, mock-jurors should have inferred that approximately $1,400,000 was owed, and this was also

specifically highlighted in one of the attorney's closing arguments. Few mock-jurors (8%) took into account the losses incurred by the company in the year preceding the fire, which amounted to one million dollars; 69% listed only the upcoming loan repayment of $400,000 due two weeks after date of the fire. The remaining 23% of the responses ranged between these two amounts.

Subjects were also asked a few questions about the procedures followed by the forensic scientists in testing the gasoline samples to assess how well they understood factors underlying the probabilistic evidence. Subjects were asked how many gasoline samples were tested. 73% of the subjects answered correctly that four samples were compared. By comparison, when asked how many components of the gases were compared in assessing their similarity, most of the subjects, 78%, were unable to answer correctly.

To assess comprehension and inferences drawn from the expert testimony subjects were asked to provide a probability rating indicating the likelihood that the gasoline samples taken from the scene of the fire and the gas tank of the car belonging to the suspected arsonist matched by chance alone. The phrasing of this question precisely reflected the language in which this probabilistic information was presented in the summary of testimony by the expert Fire Chief Olson, and again on Exhibit B. Only 13% of the subjects indicated that the probability was less than 1%, despite the fact that this factual assertion was never contested by the opposing expert. Mock-jurors' responses ranged across the entire scale, zero to 100. As many as 17% of the mock-jurors believed there was a 50% likelihood that

the gasoline samples would match "by chance alone". A second error cluster (17% of the mock-jurors) occurred at 10%, possibly because this is a multiple of the stated probability, .001.

To assess mock-jurors' understanding of the probabilistic information, they were asked to express the given probability in terms of odds as well as as a percentage. On the whole, mock-jurors who received the probabilistic information in one form were not able to answer the question seeking the information in the unfamiliar format. Many simply wrote the information down in the format in which it was originally presented, no matter what format they were prompted for. For example, when asked to state the odds, of the 63% of the subjects who did respond in the appropriate format, 50% were originally presented with the information in that desired format. Seventeen percent of the responses in the odds format were erroneous, all errors made in the same direction, i.e., the likelihood of a chance match was overstated as 1/99 or 1/100 (15%) or as 1/10 (2%).

Similarly, when asked to express the likelihood of matching gasoline samples by chance alone as a percentage, once again, 65% of the responses received were in an appropriate format, of which 50% had previously been exposed to the information presented in this fashion. Errors were more frequent, and erroneous responses varied more than did responses to the analogous odds question. Approximately 15% of the subjects believed there was a 1% probability in lieu of .1%, and 16% of the subjects stated some other incorrect percentage greater than 1%.

To assess whether mock-jurors' responses to the probabilistic scientific evidence were influenced by their attitudes towards

mathematics, responses to the mathematics attitudes questions were summed. On the whole, mock-jurors did not report very negative attitudes toward mathematics. On a 48-point scale ranging from 12 to 60 on which higher scores indicated greater anxiety, the mean score was 33. On a similar scale for confidence in mathematics, on which higher scores indicated less confidence, the mean score was 32. These scores were highly correlated ($\underline{r}$ = .87, $\underline{p}$ < .01). Accordingly, subjects' scores on these scales were added to produce a combined total mathematics anxiety score. Scores on this scale were split at the median. There were significantly more women than men in the high anxiety group (58% vs. 32%, N = 213). Chi-square (1) = 14.43, $\underline{p}$ = < .01.

A chi-square analysis of mock-jurors' verdicts showed a significant relationship between mathematics experience and mathematics anxiety. Chi-square (1, N = 213) = 24.45, $\underline{p}$ = < .01. However, there was no significant effect of math anxiety on verdict. Pearson product moment correlation coefficients were computed to explore the relationship between dependent variables and other individual difference variables, such as educational background and mathematics experience. Mock-jurors' estimates of the likelihood that chance factors alone accounted for the similarity of the gasoline samples were significantly correlated with mock-jurors' mathematics experience ($\underline{r}$ = -.22, N = 192, $\underline{p}$ = < .05), and with juror's educational background ($\underline{r}$ = -.14, N = 194, $\underline{p}$ = < .05). A Pearson product-moment correlation also revealed a significant relationship between mock-jurors' mathematics experience and culpability estimates based on the similarity of the gasoline

samples ($\underline{r}$ = .13, N = 194, $\underline{p}$ = < .05). Correlation coefficients of mock-jurors' confidence in their verdicts and either their mathematics experience or their educational background were not significant.

To further explore the relationship between individual differences and the extent to which mock-jurors perceived the scientific evidence as incriminating, a step-wise multiple regression was computed. This analysis permitted an assessment of the proportion of variance in mock-jurors' responses attributable to individual differences such as mathematics experience or mathematics anxiety, or to the independent variables.

In response to questions about issues of ultimate fact, i.e., whether the fire was the result of arson or an accident, the contribution of mock-jurors' mathematics experience to the variance in responses was not significant. In response to questions which evaluated inferences of guilt based upon the scientific evidence itself, mathematics experience accounted for a significant proportion of the variance in mock-jurors' responses. When subjects estimated the extent to which chance factors alone account for the similarity between the matching gasoline samples, mathematics experience accounted for .06 of $\underline{r}$-squared ($\underline{b}$ = -.24, $\underline{p}$ = < .01). When subjects estimated the extent to which the similarity of the gasoline samples indicated that the fire was the result of arson, mathematics experience accounted for .03 of $\underline{r}$-squared ($\underline{b}$ = .16, $\underline{p}$ = < .05). In response to a question which examined the extent to which the scientific evidence was perceived as incriminating, the two major variables contributing to variation in responses were mathematics experience, and to a lesser degree,

mathematics anxiety. These variables accounted for .02 and .01 of *r*-squared, respectively. The contribution of the first variable was statistically significant, ($p < .05$), while that of mathematics anxiety was not ($p > .05$). Notably, the three independent variables accounted for less variance in the manner in which the scientific evidence was perceived than did the mock-jurors' prior mathematics experience.

To examine the unique contribution to the variance by certain predictor variables, a descending multiple regression was computed on several of the dependent variables. Variables were entered into the regression in the order in which they might influence use of the probabilistic scientific information by the mock jurors, in light of the information-integration model discussed earlier. First, mock-jurors' attitudes towards mathematics were entered, regarded as a permanent trait which mock-jurors bring to the task. Second, the three independent variables were entered: case, linguistic form of the evidence, and presence or absence of an illustration. Finally, demographic variables were entered: gender, educational background, mathematics experience. The regression was computed on the following dependent variables: verdict, estimates of defendant's guilt, confidence in verdict and weight of expert testimony.

By means of this analysis, two variables emerged as significant. First, one of the independent variables, the presence or absence of a bar-graph to illustrate the results of the scientific tests performed on the gasoline samples accounted for .04 of *r*-squared ($b = .17$; $p < .05$). Second, one of the individual difference measures, mock-jurors' mathematics experience, accounted for more of the variance (r-squared =

.49, b =.18; $p < .05$). These analyses complemented and extended the results of the step-wise multiple regression and the Pearson product-moment correlations.

Overall, mathematics anxiety was not a useful predictor of mock-jurors' responses to probabilistic scientific evidence. There was also no support in the data for the hypothesis that jurors with high mathematics anxiety are more prone to utilize the probabilistic evidence. By comparison, jurors' prior mathematics experience was a more useful predictor of jurors' responses.

Comprehension and recall of jury instructions:

First, mock-jurors were asked to note the most important instruction from the judge. 37% of the subjects listed the burden of proof, 25% listed the importance of considering the facts in evidence. A further 8% mentioned the instruction to exclude the opening and closing statements by the attorneys in a consideration of the facts.

To determine whether mock-jurors understood the difference between direct and circumstantial evidence, and were able to apply these concepts expressed in the judge's instructions to the facts in evidence, mock-jurors were asked to cite one example each of direct and circumstantial evidence in this case. Here again, errors were common. As many as 29% of the mock-jurors erroneously believed the chromatographic gasoline analyses constituted direct evidence. Other items of circumstantial evidence, such as the presence of Avery's body at the scene of the fire, the gasoline cans, and the gasoline samples taken by the forensic scientists were also listed as direct evidence by 32% of the mock-jurors. Errors in the interpretation of what

constitutes circumstantial evidence also occurred. Only 31% of the subjects mentioned the gasoline chromatography tests as circumstantial evidence; however, other suitable responses included the increases to the amount of insurance coverage (17%) and Jackson's financial predicament (9%). A sizeable number of mock-jurors believed that hearsay evidence was circumstantial, i.e., 20% of the mock-jurors believed that testimony by Marie Williams that she overheard Avery referred to as a "torch" was circumstantial rather than direct evidence, and a further 8% listed other witness statements in response to this question.

Most Important Facts

To obtain some indication as to whether mock-jurors tended to ignore or dismiss the scientific evidence, mock-jurors were asked to note, in order of importance, three major facts which influenced their verdicts. Facts could be noted either in support of a decision that the fire was accidental or in support of a decision in favor of arson.

In all, 43% of the subjects mentioned the matching gasoline samples as one of the three major reasons for their verdict. Of this group, 22% mentioned the evidence about the matching gasoline samples first, as the most important fact. Another 17% of the mock-jurors referred to the testimony of the expert who introduced the evidence of the matching gasoline samples, without specifically mentioning the probabilistic testimony. Most of the mock-jurors in this group (83%) returned a verdict of arson. The matching gasoline sample tests were cited by 11% of the subjects as the second most important fact, and by 10% as the third most important fact. Three other facts received frequent

mention: (1) the presence of Avery at the scene of the fire and (2) Jackson's financial motive to cause the fire deliberately, and (3) the increase of the insurance coverage shortly before the date of the fire.

Reasons for Verdict

Subjects' written responses providing a reason for their verdict were analyzed to determine whether the probabilistic evidence was given prominence in justifying the decisions. Overall, 26% of the subjects mentioned the scientific evidence regarding the matching gasoline samples in response to this question. Subjects mentioned this evidence both in support of a decision favoring arson and in support of a decision rejecting arson. A few examples of responses mentioning the scientific evidence are listed below. In each case, the entire response provided by the subject to this question is listed in paragraph format:

> The chain of circumstantial evidence--company losing money and having debt. The timing of raising insurance. The testimony by Williams at the hearing of Avery. The similarity between the gas in Avery's car and on the floor of the store room.

> Marie Williams was jealous and vindictive--also do not feel the gas test conclusive. Building a potential fire hazard.

> The state did not go far enough to prove Avery was not an engineer--they implied this fact. Too much weight of prosecution on chemical test. But nobody tried to find where the gas came from.

> There was a preponderance of evidence that the owner was desperate, had tried to cover bases (i.e., insurance increases), had arranged a means to perpetrate the act, and the gasoline types and evidence were compatible to such an act.

By comparison, some examples of responses which omitted mention of the scientific evidence are:

> Marie Williams heard Avery referred to as a torch, her affair with Jackson was over for some time, so she now had not so much of her personal feelings involved, to try to hurt Jackson.
>
> If they wanted to set fire to the business it would seem they would have hidden "Avery" not introduce him to people.
>
> Independent investigator was never on scene of fire. Administrative aide did not check out background of new employee. Insurance increased Oct. 1984.
>
> Jackson was in financial trouble and the bank would not give him a new loan. He thought he would lose his business so he hired someone to set fire and collect the insurance.
>
> An old building containing machinery, flammable materials and human frailties, catching fire is easily believable. No-one actually saw anything to prove arson.

The 48 responses which contained mention of the probabilistic evidence were subjected to further qualitative analyses to determine whether the mock-jurors (a) interpreted the evidence erroneously; (b) appeared to be mesmerized by the evidence; or (c) discounted its value.

Only one response clearly fell into the first group. It contained a clear substitution error of 1% in place of .1%.:

> Difficult case to decide. Would have preferred to witness interrogation. Analysis of gasoline very important 1% from Avery's car and concrete floor important - financial difficulties entered in my thinking - fire inspector waited till following day.

A few examples of responses were found in which the scientific evidence appeared to be accorded undue weight, in approximately 6% of the responses (11/192). Although these responses cannot be classified as exemplars of the prosecutor's fallacy, because there is no

indication that there is any confusion between this evidence and the burden of proof, these responses do indicate a prosecutorial bias. Responses in this category included:

> The similarity of the gasoline from Avery's car and that on the concrete floor, together with the fire marshal's testimony were most significant. Peterson's description of Avery as a "torch" was damning. Motive was certainly there.
>
> I believe the laboratory test of gasoline. I don't see any evidence of Avery's background of competence and I question switching to a new product line in light of the company's situation. It is not probable that lawnmower gasoline was available in a sufficient amount to trigger a 5 hr. blaze. It is totally unproven that wiring was faulty. Ms. Williams and Pinkus testimony.
>
> The only hard evidence was that of the chemical similarity of the gasoline. Financial problems. Why did you present the defendant's case first rather than the plaintiff's as is the usual procedure?

A slightly higher number of responses, 9% (17/192), were classified as examples of the defense fallacy, i.e., responses which indicated that the mock-jurors gave no weight to this evidence. Examples included:

> Gasoline sample analysis inconclusive, speculative. A number of unexplored motives by persons other than Jackson. Prosecution failed to discredit Avery's qualifications. Delay by secretary (Williams) in reporting arson claim until after there was a fire.
>
> I feel that the prosecution has not presented a case "beyond reasonable doubt" in its evidence, that the circumstances of the fire were such as to a strong possibility of accidental eruption--I don't feel that an accelerant was a factor.
>
> It is suspicious, the circumstances--there are a good many why's. But it seems that fires on old buildings can start easily, and the gas samples aren't convincing--couldn't the gardener have stolen the gas and bought a bottle with the cash? My questions aren't answered--I have doubts.

> Do not believe an arsonist would be killed. Gasoline samples were not same. Understand other jealous woman.

In the balance of the responses in which the gasoline samples were mentioned (18/192) the mock-jurors appeared to give the scientific evidence more even consideration, somewhere between the extremes of overreliance and total discounting. Examples of such responses are:

> The aluminum company was in desperate need of money; the insurance was increased at an unusual date; Avery was hired to be an undesignated job and paid in advance; he died near the point of origin of the fire; gasoline found there matched that in his car.
>
> Jealousy of employee; plans late but in process created by Avery at price of $5,000 (seemed reasonable): gas could have come from another automobile; Jackson's wife may not have known how terribly he needed the money (pride of Jackson's); testimony of fire fighter; age of the building.
>
> No-one checked credentials of Avery. Is he an expert designer or not? Should have been established. He sure could not be an established arsonist. Williams not an impartial witness. Gasoline tests only prove it was siphoned out of Avery's car. Not guilty due to lacking preponderance of the evidence.
>
> The fact that the business was failing and that he raised the insurance just prior to the fire and the testimony of the fire chief that it was arson that caused the fire. The chromatography tests were quite conclusive.

While justifications such as these, provided after mock-jurors had rendered their verdicts, are indirect measures of their decision-making processes (Ericsson & Simon, 1980), they are nevertheless of value in pinpointing the types of inferences which mock-jurors draw from the evidence which they considered in the course of reaching a verdict.

CHAPTER FIVE

CONCLUSION

Twenty years ago, Kalven and Zeisel (1966) noted that the question of juror competence cannot be answered _a priori_. They conducted an empirical inquiry to determine whether jurors understand the facts presented in a case and render a verdict in accordance with the law. Based on post-trial interviews of actual jurors and judges, they concluded that, by and large, jurors do return verdicts in line with the evidence presented. Since the publication of their seminal work in this area, the nature of evidence presented in trials has become increasingly scientific, and methods to evaluate jury behavior have become more sophisticated. While questions regarding juror competency are still being posed, the issues underlying these questions are new, as are the methods applied to answer these questions.

The backdrop for these studies was a series of _a priori_ judicial opinions that recur in judicial opinions concerning jurors' underuse, overuse and misuse of probabilistic evidence. Following the suggestion by Lempert (1981) that simulation studies are best-suited to test what jurors do poorly, what they do well, and to test methods of presenting evidence to jurors, two pencil-and-paper simulation studies were conducted to address the following issues:

Do mock-jurors overbelieve scientific evidence?

These studies provided no support for the judicial position that probabilistic information should be excluded from jury trials on grounds that mock-jurors are mesmerized by the apparent precision of

science. In Study One, in which the incriminating value of the probabilistic evidence was varied from .1% to 10% (odds of a random match ranged from 1/1000 to 1/10) only 23% of the student mock-jurors who considered the probabilistic evidence found it adequate to convict the defendant. In Study Two in which fairly incriminating probabilistic evidence (odds of random match were 1/1000) was introduced, 53% of the subjects did not find the evidence adequately persuasive to return a verdict of arson. In the second study, mock-jurors provided written statements outlining the major reasons for their verdicts. While no claim is made that the written responses represent anything more than post-hoc rationalizations for the mock-jurors' verdicts, the responses in support of a verdict in favor of arson contained mention of several facts in addition to the scientific evidence. Less than 40% of the mock-jurors rated either the scientific evidence or the expert who presented it as the most important factual information in the case.

In the first study, mock-jurors' ratings of the reliability of blood tests, statistics and eyewitness testimony showed that blood tests were rated most favorably. Nevertheless, the mean was 79%, lower than reliability rates reported by forensic scientists. Accordingly, concern that probabilistic evidence will eclipse other evidence may not be well-founded. Mock-jurors appear to be somewhat cautious and skeptical about scientific evidence.

Austin (1982) theorized that jurors' skepticism about scientific evidence was based on unfamiliarity with the subject matter. Another reason jurors may be cautious about probabilistic evidence is that they

discount the value of circumstantial evidence despite instructions from the judge that they may consider this evidence in reaching a verdict (Pryor et al., 1980).

Do mock-jurors overbelieve expert witnesses?

The second reason the courts have regarded probabilisitic scientific evidence as prejudicial in jury trials is that they doubt jurors' abilities to evaluate expert testimony, and presuppose that jurors will overbelieve expert witnesses. This issue was examined in Study Two.

Without reference to any particular case, when mock-jurors were asked to rate the credibility of testimony of expert witnesses as opposed to lay witnesses, they accorded significantly more weight to the former. However, when asked to rate the credibility of the six witnesses whose testimony they considered in rendering a verdict in Study Two, mock-jurors accorded the most weight to a lay witness, the banker, and not to the experts. This finding illustrated that in in this case, mock-jurors were not overwhelmed by the expert opinions. If a laywitness and an expert were to testify about the same subject matter, however, jurors might find the expert more credible. Mock-jurors were able to distinguish between credibility assessments and the probative value of the testimony in light of the issues to be decided in the case, as was shown by the fact that testimony from the witness rated most credible did not receive the more weight from mock-jurors in reaching a verdict.

This study was also informative about mock-juror conduct when faced with a "battle of experts". Contrary to the hypothesis that jurors

exposed to two experts whose views are contradictory will ignore both experts, mock-jurors in Study Two appeared to evaluate the testimony of the experts carefully in deciding to which to accord the most weight. Factors such as whether the expert's opinion was based on an on-site inspection were cited by mock-jurors as reasons for siding with one expert over the other. Estimates of the relative weight accorded to the expert witnesses showed that, despite the "battle of the experts", expert testimony received more weight than did the testimony of lay witnesses. Given the fact that the lay witnesses (with the exception of the banker) each had various interest biases, and that there were no eyewitnesses to the alleged acts of arson, this evaluation was reasonable. The concern expressed by Saks and Van Duizend (1983) that jurors may not believe experts at all is not supported by this research. The expert who presented probabilistic evidence was rated as significantly more credible than the expert who presented no probabilistic evidence. This finding replicated that of Faigman (1983), who found that a physician who presented probabilistic information received considerably more weight than a statistician who presented no probabilistic evidence, but who explained how to use Bayes' Theorem in reaching a verdict.

Do mock-jurors accord appropriate weight to scientific evidence?

Saks and Kidd (1980) hypothesized that jurors would underuse probabilistic information. The first problem in addressing this issue is the selection of an appropriate group of comparators. Study One indicated that in comparison with Bayesian norms, student mock-jurors tend to underuse scientific information. However, the Bayesian model

has not proved representative of the way in which people reason in general (Beach, Mitchell, Deaton & Prothero, 1978). Comparisons with Bayesian norms were not possible in Study Two because there was no control group to provide a prior estimate of guilt in absence of scientific evidence. In the first study, the probabilistic evidence received more weight than any other item in evidence, and subjects who received the most incriminating probability (1/100) accorded the evidence the most weight, while subjects who received the least incriminating evidence accorded it the least weight. This results tend to show appropriate use of the evidence, even if the degree of the use is lower than that indicated by Bayesian standards.

However, if mock-jurors' uses of probabilistic evidence are compared with those of judges, it is more difficult to make the argument that mock-jurors underuse this information. A review of judicial uses of statistical evidence showed a tendency among judges to ignore or undervalue the weight of survey evidence (Jacoby, 1985). Mock-jurors appear to exhibit the same tendency. Mock-jurors' written answers stating reasons for their verdict in Study Two provided some evidence of discounting. In addition to discounting that may occur because probabilistic evidence is circumstantial, mock-jurors may tend to discount the value of probabilistic evidence because they are naive about probability or because they fail to understand the probabilistic evidence.

Do mock-jurors understand probabilistic evidence?

The comprehension measures used in these studies were instrumental in pinpointing some common inferential errors in mock-jurors'

understanding and application of probabilistic evidence in reaching a verdict. For example, when mock-jurors were asked to provide a probability estimate of a random match between the gasoline samples tested in Study Two, despite the fact that this information was provided twice in the trial materials and was not contested by the opposing expert, few responses were accurate. Most probability estimates were very high, diminishing or discounting the value of the evidence. The fact that there were error clusters at 10% and 50% may indicate both errors of comprehension and errors of application rather than deliberate discounting of this evidence. For example, one possible explanation for the cluster of responses at the 50% mark (17%) is that a number of mock-jurors are naive about probability, and believe the word chance implies a probability of .5, not realizing that this is simply one instantiation of random probability. The cluster of errors at 10% (17%) may indicate that mock-jurors are prone to make errors with multiples of ten when dealing with fractions or decimals, i.e., a computational error of application may have produced this result.

The finding that most mock-jurors had little facility with probabilities, and were unable to convert odds to percentages and vice versa, highlights the importance of mathematics experience. While Faigman (1983) reported no effect of mathematics experience on verdict, he did report enhanced comprehension of the Bayesian presentation by students with experience in statistics. In Study Two, mathematics experience emerged as the best predictor of mock-jurors' responses to probabilistic evidence, accounting, however, for only a small

proportion of the variance in their answers. A few mock-jurors admitted they did not understand the evidence. Only a very small percentage of the mock-jurors appeared to be susceptible to the prosecutor's fallacy. Responses of a few others indicated some susceptibility to the defense fallacy. Further investigations of differences in the skills and reasoning of individuals familiar with probability theory and those naive or lacking experience in this area is needed, as is research on ways to remedy the problem. For example, simple inoculation may prove effective, such as telling jurors that the probability of a random match does not necessarily mean there is a 50-50 chance of a match.

Influence of the form of probabilistic evidence

The results of variations in the form of the probabilistic evidence are less easy to interpret. Thompson (1984) reported that, when the probabilistic information was presented in the form of percentages, mock-jurors were more likely to ignore the evidence, and instances of the defense fallacy increased. By comparison, when the evidence was presented as a discrimination probability in his studies, more prosecutorial errors occurred. This pattern of results was replicated in Study Two. The reasons for this finding are more elusive. The use of percentages as opposed to odds had a similar effect on mock-jurors' responses as did the use of a visual aid--estimates of defendant's culpability tended to decrease.

Do mock-jurors confuse the burden of proof with probabilistic evidence?

Very few mock-jurors or students appeared to be susceptible to the

prosecutor's fallacy; thus neither study produced support for the judicial notion that jurors will confuse the burden of proof with the probabilistic evidence. The probabilities used in these studies ranged from 10% to .1%. The fact that the number of findings of arson in the second study was invariant when a different burden of proof was introduced was instrumental in clarifying that confusion of probabilistic evidence with the burden of proof should not be a major cause of concern. In an actual trial, given the opportunity for discussion of the burden of proof during deliberation, the likelihood of this error may diminish.

By far the greater concern to judges should be jurors' propensity to misunderstand or misapply the burden of proof itself, independent of any probabilistic testimony. In Study Two, in general, no significant differences arose as a result of variations in the burden of proof. In other words, responses of mock-jurors' who received the civil version of the trial materials and a "preponderance of the evidence" standard of proof, were undifferentiated from responses of mock-jurors who received the criminal version of the trial materials and a standard of proof "beyond a reasonable doubt".

In Study One, the range of estimates of guilt which were regarded as adequate for a conviction beyond a reasonable doubt was between 50% and 99%, with most below 90%. The fact that variations in the burden of proof applied in the second study produced no corresponding change in the number of findings of arson indicates that mock-jurors may modify their interpretations of the burden of proof in accordance with the evidence. More mock-jurors mentioned the burden of proof when they

found the evidence inadequate to support a finding of arson. This may merely be a post-hoc rationalization, after a verdict is reached, and not an indication that the burden of proof was employed during the task of reaching a verdict. Mock-jurors who returned a verdict of arson may have modified their interpretations of the burden of proof to suit the outcome they favored, if for example, they were reasoning in a verdict-driven as opposed to an evidence-driven manner, as has been posited by Hastie et al. (1984). Because the written verdict justifications do not constitute direct evidence of mock-jurors' decision-making processes, these responses are of limited value in assessing this issue. Perhaps jury instructions which incorporate a quantitative explanation will assist jurors in properly applying the burden of proof to the evidence, despite concern by some legal theorists that an increase in the efficiency of juries may upset the delicate balance of justice (Loh, 1985).

Limitations of the studies

Two major procedural differences distinguish procedures used in these studies and those which apply in the case of an actual trial in which probabilistic evidence is introduced. First, these mock-jurors did not have an opportunity to see and hear the witnesses, for they read summaries of trial materials in which the evidence was either uncontroverted (Study One) or controverted (Study Two). Second, none of the mock-jurors in these studies participated in deliberation. These procedural differences limit the generalizability of these findings to what may take place in an actual trial. Collectively, jurors may correct each others' inferential errors, and differences in

the interpretation of the burden of proof may be minimized as a result of deliberation.

In a review of peoples' everyday inductive reasoning abilities, Nisbett, Krantz, Jepson and Fong (1982) noted that human reasoning abilities change with new cultural inventions:

> Virtually every educated person today employs essentially statistical reasoning in some domains, for example, sports and weather, and has a statistical understanding of randomizing devices like cards and dice that is very different from premodern conceptions. (p. 457.)

They envisaged a time in which statistical heuristics would be incorporated into most people's everyday reasoning. They concluded their review by stating that for the present, the important questions to be solved are (a) teaching techniques that will prove most effective to familiarize people with probabilistic models, and (b) what the best inductive principles and methods to teach are. The demand for jurors who can understand and apply probabilistic concepts in modern trials has created an everyday domain in which the need for statistical heuristics is immediate. Thse studies represent a first step in attempting to answer the questions posed by Nisbett et al. in the domain of law and psychology.

REFERENCES

Abney, D. L. (1986). Expert testimony and eyewitness identification. Case and Comment, 91, 26-38.

Arnold, M. S. (1980). A historical inquiry into the right to trial by jury in complex civil litigation. University of Pennsylvania Law Review, 128, 829-848.

Aschenbrenner, K. M. (1978). Single-peaked risk preferences and their dependability on the gambles' presentation mode. Journal of Experimental Psychology, 4(3), 513-520.

Austin, A. D. (1982). Jury perceptions on advocacy: A case study. Litigation, 8(4), 15-17.

Austin, A. D. (1984). Complex litigation confronts the jury system: A case study. Frederick, Maryland: University Publications of America, Inc.

Barclay, S., Beach, L. R., & Braithwaite, W. P. (1971). Normative models in the study of cognition. Organizational Behavior and Human Performance, 6, 389-413.

Barefoot v. Estelle, 103 S. Ct. 3383, 3398, n. 7 (1983).

Bar-Hillel, M. (1984). Probabilistic analysis in the legal fact-finding process. Acta Psychologia, 48, 71-86.

Beach, L. R., Mitchell, T. R., Deaton, M. D., & Prothero, J. (1978). Information relevance, context and source credibility in the revision of options. Organizational Behavior and Human Performance, 21, 1-16.

Beskind, D. H., Bocchino, A. J., Ordover, A. P., & Seckinger, J. H. (1983). Problems and cases in trial advocacy (Vol. 2, cases). St. Paul, MN: The National Institute for Trial Advocacy, Inc.

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. Journal of Personality and Social Psychology, 45(6), 1185-1195.

Bradley, J. V. (1981). Overconfidence in ignorant experts. Bulletin of the Psychonomic Society, 17(2), 82-84.

Braun, L. J. (1982). Quantitative analysis and the law: Probability theory as a tool of evidence in criminal trials. Utah Law Review, 4, 41-87.

Brilmayer, L., & Kornhauser, L. (1978). Review: Quantitative methods and legal decisions. The University of Chicago Law Review, 6, 116-153.

Broeder, D. W. (1959). The University of Chicago jury project. Nebraska Law Review, 32, 744-760.

Brown, R. A. (1943). Fact and Law in Judicial Review. Harvard Law Review, 56, 899-928.

Buchanan, R. W., Pryor, B., Taylor, K. P., & Strawn, D. U. (1978). Legal communication: An investigation of juror comprehension of pattern instructions. Communication Quarterly, 26, 31-35.

Burger, W. E. (1980). Can juries cope with multi-month trials? The American Journal of Trial Advocacy, 3(3), 449-458.

Capon, N., & Kuhn, D. (1979). Logical reasoning in the supermarket: Adult females' use of a proportional reasoning strategy in an everyday context. Developmental Psychology, 15, 450-452.

Carr, A. T. (1980). People, probabilities and the law. In S. L. Loyd-Bostock (Ed.), Law and psychology (pp. 157-165). Oxford, England: SSRC Center for Socio-Legal Studies.

Casper, J. D., Benedict, K., & Kelly, J. R. (1985). Cognitions, attitudes and decision-making in search and seizure cases. Paper presented at the Annual Meeting of the Law and Society Association, San Diego, California, June 6-9, 1985.

Cavoukian, A., & Heslegrave, R. J. (1980). The admissibility of polygraph evidence in court: Some empirical findings. Law and Human Behavior, 4, 117-131.

Chaiken, S., & Eagly, A. H. (1976). Communication modality as a determinant of message persuasiveness and message comprehensibility. Journal of Personality and Social Psychology, 34, 605-614.

Charrow, R. P., & Smith, R. L. (1976). A conversation about "A conversation about Collins". The Georgetown Law Journal, 64, 669-678.

Christensen-Szalanski, J. J. (1980). A further examination of the selection of problem-solving strategies: The effects of deadlines and analytic aptitudes. Organizational Behavior and Human Performance, 25, 107-122.

Clark, W. F. (1969). Scientific evidence. In B. J. George, Jr., & I. Cohen (Eds.), The prosecutor's sourcebook, (pp. 369-376). NY: The Practising Law Institute.

Clute, P. S. (1984). Mathematics anxiety, instructional method, and achievement in a survey course in college mathematics. Journal for Research in Mathematics Education, 15, 50-58.

Coates, D., & Penrod, S. (1980). Social psychology and the emergence of disputes. Law and Society Review, 15(3-4), 655-680.

Constanzo, M., Archer, D., Arsonson, E., & Pettigrew, T. (1986). Energy conservation action: The difficult part from information to action. American Psychologist, 21(25), 521-528.

Coombs, C. H., Donnell, M. L., & Kirk, D. B. (1978). An experimental study of risk preferences on lotteries. Journal of Experimental Psychology: Human Perception and Performance, 4, 497-512.

Curtis, W. C., & Wilson, L. W. (1979). The use of statistics and statisticians in the litigation process. Jurimetics Journal, 20, 109-120.

Dane, F. C. (1985). In search of reasonable doubt: A systematic examination of selected quantification approaches. Law and Human Behavior, 9, 141-158.

Dellarosa, D., & Bourne, L. E. (1984). Decisions and memory: Differential retrievability of consistent and contradictory evidence. Journal of Verbal Learning and Verbal Behavior, 23, 669-682.

Devine, P. G., & Ostrom T. M. (1985). Cognitive mediation of inconsistency discounting. Journal of Personality and Social Psychology. 49, 5-21.

Edwards, W. (1962). Subjective probabilities inferred from decisions. Psychological Review, 69(2), 109-135.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 32, 52-88.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87(3), 215-251.

Faigman, D. L. (1983). Bayes' Theorem in the trial process: Instructing juries on the value of probabalistic evidence. Paper presented at the meeting of the American Psychological Association, Anaheim, CA.

Fennema, E. (1976). Fennema-Sherman Mathematics Attitude Scales. JSAS Catalog of Selected Documents in Psychology, 6, Ms. No. 1225, 31.

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. American Educational Research Journal, 14,(1), 51-77.

Finfrock, W. P., & Spradlin, B. C. (1978). How to organize and present statistical evidence. The Practical Lawyer, 24 (4), 67-76.

Finkelstein, M. O., & Fairley, W. B. (1970). A Bayesian approach to identification evidence. Harvard Law Review, 83(3), 489-517.

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. Memory and Cognition, 12, 402-410.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. Psychological Review, 90, 239-260.

Fischoff, B., & Lichtenstein, S. (1976). A little learning ...: Confidence in multicue judgment. In R. Nickerson (Ed.), Attention and performance, VIII (pp. 552-564). Hillsdale, NJ: Erlbaum.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. Journal of Experimental Psychology: Human Perceptions and Performance, 4, 330-334.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

George, B. J., Jr. (1981). Statistical problems relating to scientific evidence. In E. J. Imwinkelried (Ed.), Scientific and expert evidence (pp. 105-138). NY: Practicing Law Institute.

Giannelli, P. C. (1980). The admissibility of novel scientific evidence: Frye v. United States, a half-century later. Columbia Law Review, 80, 1197-1246.

Giannelli, P. C. (1983). [Symposium remarks.] In W. A. Thomas (Ed.), Symposium on science and rules of evidence. Federal Rules Decisions, 99, 187-234.

Gick, M. (1983). Diagrams as retrieval cues for analogous problems in memory. Paper presented at the conference on inductive reasoning, Rosario, Orcas Island, WA.

Graham, M. (1983). [Symposium remarks.] In W. A. Thomas (Ed.), Symposium on science and rules of evidence. Federal Rules Decisions, 99, 187-234.

Gray, M. W. (1986). Legal perspectives on sex equity in faculty employment. Journal of Social Issues, 41(4), 121-134.

Greene, E., Schooler, J. W., & Loftus, E. F. (1985). Expert psychological testimony. In S. M. Kassin & L. S. Wrightsman (Eds.), The psychology of evidence and trial procedure (pp. 201-252). Beverly Hills, CA: Sage Publications, Inc.

Hallock, M. M. (1977). The numbers game: The use and misuse of statistics in civil rights litigation. Villanova Law Review, 23, 5-34.

Hastie, R. (1983). Social inference. Annual Review of Psychology, 34, 511-542.

Hastie, R., Penrod, S. D., & Pennington, N. (1983). What goes on in a jury deliberation. American Bar Association Journal, 69, 1848-1853.

Hastie, R., Penrod, S. D., & Pennington, N. (1984). Inside the jury. Cambridge: Harvard University Press.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Hedrick, E. R. (1928). The reality of mathematical processes. Yearbook 3: National Council of Teachers of Mathematics, 35, 40-41.

Imwinkelried, E. J. (1981). A new era in the evolution of scientific evidence: A primer on evaluating the weight of scientific evidence. William and Mary Law Review, 23, 261-291.

Imwinkelried, E. J. (1982). The standard for admitting scientific evidence: A critique from the perspective of juror psychology, Villanova Law Review, 28, 554-571.

Imwinkelried, E. J. (1983). Methods of attacking scientific evidence. Charlottesville, VA: Michie.

In re Boise Cascade Securities Litigation, 420 F.Supp. 99, 104 (W.D. Wash. 1978).

Jacoby, J. (1985). Survey and field experimental evidence. In S. M. Kassin & L. S. Wrightsman (Eds.), The psychology of evidence and trial procedure (pp. 175-200). Beverly Hills, CA: Sage Publications, Inc.

Jaffee, L. R., (1979). Comment on the judicial use of HLA paternity test results and other statistical evidence: A response to Terasaki. Journal of Family Law, 17, 457-484.

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. The Behavioral and Brain Sciences, 8, 193-263.

Jonakait, R. N. (1983). When blood is their argument: Probabilities in criminal cases, genetic markers, and, once again, Bayes' Theorem. University of Illinois Law Review, 2, 369-421.

Jones v. State, 208 S.E.2d 850, 853, 232 Ga. 762 (1974).

Kagehiro, D. K., & Stanton, C. W. (1985). Legal vs. quantified definitions of standards of proof. Law and Human Behavior, 9(2), 159-178.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80(4), 237-251.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

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

Kaplan, M. F., & Scherschíng, C. (1980). Reducing juror bias: An experimental approach. In P. Lipsitt & B. D. Sales (Eds.), New directions in psycholegal research (pp. 149-170). New York: Van Nostrand Reinhold.

Kerr, N. L. (1978). Severity of prescribed penalty and mock jurors' verdicts. Journal of Personality and Social Psychology, 36(12), 1431-1442.

Kruglanski, A. W., Friedland, N., & Farkash, E. (1984). Lay persons' sensitivity to statistical information: The case of high perceived applicability. Journal of Personality and Social Psychology, 46, 503-518.

Lempert, R. O. (1981). Civil juries and complex cases: Let's not rush to judgment. Michigan Law Review, 80, 68-131.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases, (pp. 307-332). New York: Cambridge University Press.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science, 9, 563-564.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. Journal of Experimental Psychology, 89, 46-55.

Loftus, E. F. (1980). Psychological aspects of courtroom testimony, Annals of the New York Academy of Sciences, 347, 27-37.

Loh, W. D. (1979). Some uses and limits of statistical and social science in the judicial process. In L. E. Abt & I. R. Stuart (Eds.), Social psychology and discretionary law (pp. 18-43). New York: Van Nostrand.

Loh, W. D. (1984). Social research in the judicial process: Cases, readings and text. New York: Russell Sage Foundation.

Loh, W. D. (1985). The evidence and trial procedure: The law, social policy, and psychological research. In S. M. Kassin & L. S. Wrightsman (Eds.), The psychology of evidence and trial procedure (pp. 13-39). Beverly Hills, CA: Sage Publications, Inc.

Markwart, A., & Lynch, B. E. (1979). The effect of polygraph evidence on mock-juror decision-making. Journal of Police Science and Administration, 7, 324-332.

Melani v. Board of Higher education, No. 73-5434 (S.D.N.Y. March 17, 1983).

Meyers, J. L., Hansen, R. S., Robinson, R. C., & McCann, J. (1983). The role of explanation in learning elementary probability. Journal of Educational Psychology, 75(3), 374-381.

Mode, E. B. (1963). Probability and criminalistics. Journal of the American Statistical Association, 58, 628-640.

Moenssens, A. (1983). [Symposium remarks.] In W. A. Thomas (Ed.), Symposium on science and the rules of evidence. Federal Rules Decisions, 99, 187-234.

Monahan, J., & Loftus, E. F. (1983). The psychology of law. Annual Review of Psychology, 441-475.

Monahan, J., & Walker, L. (1985). Social science in law: Cases and materials. Mineola, NY: The Foundation Press, Inc.

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

National Center for State Courts. (1980). Study to investigate the use of scientific evidence. National Center for State Courts Report, 7, 1.

New Jersey v. Cavallo, 88 N.J. 508, 433 A.2d 1020 (1982).

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

Nisbett, R. E., Krantz, D. H., Jepson, C., & Fong, G. T. (1980). Improving inductive inferences. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases, (pp. 445-459). New York: Cambridge University Press.

Pancerz, C. E. (1983). Statistics in the law: Potential problems in the presentation of statistical evidence. Washington and Lee Law Review, 40, 313-339.

Panel on Statistical Assessments as Evidence, (1985). The Evolving Role of Statistical Assessments as Evidence in the Courts. Report prepared for the National Committee on Statistics and the Committee on Research on Law Enforcement and the Administration of Justice.

Paul, D. J., Nibbelink, W. H., & Hoover, H. D. (1986). The effects of adjusting readability on the difficulty of mathematics story problems. Journal for Research in Mathematics Education, 17(3), 163-171.

Payne, J. W. (1982). Contingent decision behavior. Psychological Bulletin, 92, 382-402.

People v. Collins, 68 Cal.2d 319, 325-327, 438 P.2d 33, 38-39 (1968).

People v. Lindsey, 149 Cal. Rptr. 47, 2 A.L.R. 4th 485, (1978).

People v. Woodward, No. 108551 (Cal. Sup. Crt., San Mateo County, July 7, 1964).

Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. Journal of Personality and Social Psychology, 37(10), 1915-1926.

Presseisen v. Swarthmore College, 582 F.2d 1274 (3d Cir. 1978).

Pryor, B. K., Buchanan, R. W., & Strawn, D. U. (1980). An affective-cognitive consistency explanation for comprehension of standard jury instructions. Communication Monographs, 47, 69-76.

Rasinski, K. A., Crocker, J., & Hastie, R. (1985). Another look at sex stereotypes and social perceiver's use of subjective probabilities. Journal of Personality and Social Psychology, 49(2), 317-326.

Reed v. State, 391 A.2d 364, 370, 283 Md. 374, 385 (1978).

Rothbart, M., Fulero, S., Jensen, C., Howard, J. & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. Journal of Experimental Social Psychology, 14, 237-255.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education (pp. 33-58). Hillsdale, NJ: Erlbaum.

Saks, M. J., & Kidd, R. F. (1980). Human information processing and adjudication: Trial by heuristics. Law & Society Review, 15(1), 123-160.

Saks, M. J., & Van Duizend, R. (1983). The use of scientific evidence in litigation. Washington, D.C.: National Center for State Courts.

Seaver, D. A., von Winterfeldt, D., & Edwards, W. (1978). Eliciting subjective probability distributions on continuous variables. Organizational Behavior and Human Performance, 21, 379-391.

Severance, L. J., & Loftus, E. F. (1982). Improving the ability of jurors to comprehend and apply criminal jury instructions. Law and Society Review, 17, 153-197.

Shaklee, H., & Hall, L. (1983). Methods of assessing strategies for covariation between events. Journal of Educational Psychology, 75(4), 583-594.

Shanks v. State, 45 A.2d 85, 85-90, 185 Md. 437 (1945).

Simon, R. J. (1970). "Beyond a reasonable doubt": An experimental attempt at quantification. 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.

Slovic, P., Fischhoff, B., & Lichtenstein, L. (1976). Cognitive process and societal risk taking. In J. B. Carroll & J. W. Payne (Eds.), Cognition and social change. Potomac, MD: Erlbaum.

Sperlich, P. W. (1985). The evidence on evidence: Science and law and conflict and cooperation. In S. M. Kassin & L. S. Wrightsman (Eds.), The psychology of evidence and trial procedure (pp. 325-361). Beverly Hills: Sage Publications, Inc.

State v. Carlson, 267 N.W.2d 170, 176 (Minn. 1976).

State v. Cary, 49 N.J. 343, 352, 230 A.2d 384 (1967).

State v. Fulton, 263 S.E.2d 608 (1980).

State v. Garrison, 120 Az. 255, 585 P.2d 563 (1978).

State v. Saldana, 324 N.W. 2d (S.Ct Minn. 1982).

Stogsdill v. State, 552 S.W.2d 481 (Tex. Crim. App. 1977).

Straf, M. (1983). [Symposium remarks.] In W. A. Thomas (Ed.), Symposium on science and the rules of evidence. Federal Rules Decisions, 99, 187-234.

Strawn, D. U., & Munsterman, G. T. (1979). Helping juries handle complex cases. Judicature, 65, 444-447.

Sutton-Barbere, G., Teitelbaum, L. E., & Johnson, P. J. (1986, March). Burden of proof. Paper presented at the meeting of Div. 41 of the American Psychological Association, Tucson, AZ.

Tawshunsky, A. (1983). Admissibility of mathematical evidence in criminal law trials. American Criminal Law Review, 21, 29, 55-79.

Taylor, S. E., & Thompson, S. (1982). Stalking the elusive vividness effect. Psychological Review, 89, 155-181.

Thompson, W. C. (1984). Judgmental bias in reaction to mathematical evidence. Unpublished manuscript.

Thompson, W. C. (1985). Fallacious interpretations of statistical evidence in criminal trials. Unpublished manuscript.

Thompson, W. C. (1986, March). Mathematical evidence in criminal trials: Improving the probability of justice. Paper presented at the meeting of Div. 41 of the American Psychological Association, Tucson, AZ.

Thomas, E. A. C., & Hogue, A. (1976). Apparent weight of evidence, decision criteria, and confidence ratings in juror decision making. Psychological Review, 83(6), 442-465.

Thomas, W. A., (Ed.) (1983). Symposium on science and the rules of evidence. Federal Rules Decisions, 99, 187-234.

Time, (1981, July). The Case of the Green Carpet, p. 12.

Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. Harvard Law Review, 84(6), 1329-1393.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological Review, 79(4), 281-299.

Tversky, A., & Kahneman, D. (1971). Belief in the "law of small numbers." Psychological Bulletin, 76, 105-110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases, (pp. 117-128). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90(4), 293-315.

United States v. Addison, 498 F.2d 741, 744 (D.C. Cir. 1974).

United States v. Amaral, 488 F.2d 1148, 1152 (9th Cir. 1973).

United States v. Buck, 449 F.2d 262 (10th Cir. 1971).

United States v. Collins, 395 F.Supp. 629, 637 (M.D. Pa. 1975).

United States ex rel. DiGiacomo v. Franzen, 680 F.2d 515 (5th Cir. 1982).

United States v. Hulen, 3 M.J. 275 (CMA 1977).

United States v. Massey, 594 F.2d 676 (8th Cir. 1979).

United States v. Torniero, 735 F.2d 725, 734 (2d Cir. 1984).

Wainer, H. (1980). A test of graphicacy in children. Applied Psychological Measurement, 4(3), 331-330.

Wasserman, D. T., & Robinson, J. N. (1979). Extra-legal influences, group processes and jury decision-making: A psychological perspective. North Carolina Central Law Journal, 96-157.

Wehmhoefer, R. A. (1985). Statistics in litigation: Practical applications for lawyers. Colorado Springs, CO: Shepard's/McGraw-Hill.

Wells, G. L., & Loftus, E. F. (1983). Advances in the psychology of eyewitness testimony. Cambridge, MA: Cambridge University Press.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions and the use of evidence. Journal of Applied Psychology, 59(5), 555-561.

Williams, G. (1979). The mathematics of proof. Criminal Law Review, 297, 341-354.

Zadny, J., & Gerard, H. B. (1974). Attributed intentions and informal selectivity. Journal of Experimental Social Psychology, 10, 34-52.

Zeisel, H., & Diamond, S. S. (1974). The effect of peremptory challenges on jury and verdict: An experiment in a federal district court. Stanford Law Review, 30, 491-532.

APPENDIX A

Materials for Study One

Please read the following passage carefully:

Greg Johnson and Bill Taylor are business partners and long-time friends who have spent most of their lives in Boise, Idaho, a city of approximately 100,000 inhabitants. On September 23rd, 1983, Bill, who is generally very punctual, doesn't arrive at the office. Bill and Greg had scheduled an important business conference that day. Greg calls Bill's home phone number, but no-one answers. Bill lives alone in a suburban apartment. Greg sends his secretary to Bill's apartment to check on Bill.

Upon arriving at the apartment, the secretary finds Mr. Taylor slain, and calls the police. Forensic experts arrive and take blood samples from the hallway floor and from the victim. Later that day, while interviewing Greg Johnson to learn all he can about the victim, a police detective observes that Mr. Johnson has some scratches on his face. He notes that the scratches could have been received during a struggle or in a minor bicycle accident in which Greg was involved earlier that week.

The analysis of blood samples reveals that Bill Taylor has type O blood. The court orders Greg Johnson to submit to a blood test, and he and is found to have blood type A2B. The sample of blood taken from the apartment hallway is also type A2B. Population data reveal that blood type A2B occurs in one per cent (.01%) of the population of Boise, Idaho.

QUESTION ONE:

For each fact listed, record your response by marking the scale to indicate a value between -100 and +100 which corresponds with the manner in which you interpret that fact in evidence. You must record a value for each fact listed. Refer to the following table in recording your answers:

-100 = definite proof Greg Johnson is NOT GUILTY
-50 = more likely he is NOT GUILTY than guilty
0 = shows neither guilt nor absence of guilt
50 = more likely he is GUILTY than not guilty
100 = definite proof Greg Johnson is GUILTY

a. The grand jury indicted Greg Johnson for the murder of Bill Taylor.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

b. Bill and Greg spent most of their lives in Boise, Idaho.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

c. Boise is a city of approximately 100,000 inhabitants.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

d. Bill and Greg were business partners.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

e. Bill and Greg were long-time friends.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

f. Bill and Greg had scheduled an important business conference on September 23rd, 1983.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

g. Greg had a minor bicycle accident earlier that week.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

h. Bill was generally very punctual.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

i. Bill did not arrive at the office on September 23rd, 1983.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

j. Bill lived alone in an apartment in the suburbs of Boise.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

k. Greg tried calling Bill at home.

-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

l. Greg sent his secretary to check on Bill.

-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

m. The secretary found Bill slain.

-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

n. Greg had scratches on his face on September 23rd, 1983.

-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

o. The police detective suspected Greg knew more about Bill's murder than Greg would admit.

-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

p. Forensic experts analysed the blood samples taken from the apartment and the victim.

-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

q. Bill's blood type differed from samples of blood found in the hallway of his apartment.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

r. Blood type A2B occurs in .01% of the population of Boise.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

s. The court ordered Greg to have a blood test.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

t. Greg's blood type matched the samples of blood taken from Bill's apartment hallway.

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
-100 -80 -60 -40 -20 0 20 40 60 80 100
not guilty guilty

QUESTION TWO:

You have 100 points which must be divided among the twenty facts listed on the next page. Assign the most points to the facts which most strongly indicate that Greg Johnson is GUILTY of murdering Bill Taylor. You do not have to assign points to every fact on the list. The total number of points assigned must add up to 100.

Facts	Points
The grand jury indicted Greg Johnson for the murder of Bill Taylor.	____
Bill and Greg spent most of their lives in Boise, Idaho.	____
Boise is a city of approximately 100,000 inhabitants.	____
Bill and Greg were business partners.	____
Bill and Greg were long-time friends.	____
Bill and Greg had scheduled an important business conference on September 23rd, 1983.	____
Greg had a minor bicycle accident the previous week.	____
Bill was generally very punctual.	____
Bill did not arrive at the office on September 23rd, 1983.	____
Bill lived alone in an apartment in the suburbs of Boise.	____
Greg tried calling Bill at home.	____
Greg sent his secretary to check on Bill.	____
The secretary found Bill slain.	____
Greg had scratches on his face on September 23rd, 1983.	____
The police detective suspected Greg knew more about Bill's murder than Greg would admit.	____
Forensic experts analysed the blood samples taken from the apartment and the victim.	____
Bill's blood type differed from samples of blood found in the hallway of his apartment.	____
Blood type A2B occurs in .01% of the population of Boise.	____
The court ordered Greg to have a blood test.	____
Greg's blood type matched the samples of blood taken from Bill's apartment hallway.	____
TOTAL:	100

QUESTION THREE:

You have 100 points which must be divided among the twenty facts listed on the following page. Assign the most points to facts which most strongly indicate that Greg Johnson is NOT GUILTY of murdering Bill Taylor. You do not have to assign points to every fact on the list. The total number of points assigned must add up to 100.

Facts	Points
The grand jury indicted Greg Johnson for the murder of Bill Taylor.	____
Bill and Greg spent most of their lives in Boise, Idaho.	____
Boise is a city of approximately 100,000 inhabitants.	____
Bill and Greg were business partners.	____
Bill and Greg were long-time friends.	____
Bill and Greg had scheduled an important business conference on September 23rd, 1983.	____
Greg had a minor bicycle accident the previous week.	____
Bill was generally very punctual.	____
Bill did not arrive at the office on September 23rd, 1983.	____
Bill lived alone in an apartment in the suburbs of Boise.	____
Greg tried calling Bill at home.	____
Greg sent his secretary to check on Bill.	____
The secretary found Bill slain.	____
Greg had scratches on his face on September 23rd, 1983.	____
The police detective suspected Greg knew more about Bill's murder than Greg would admit.	____
Forensic experts analysed the blood samples taken from the apartment and the victim.	____
Bill's blood type differed from samples of blood found in the hallway of his apartment.	____
Blood type A2B occurs in .01% of the population of Boise.	____
The court ordered Greg to have a blood test.	____
Greg's blood type matched the samples of blood taken from Bill's apartment hallway.	____
TOTAL:	100

QUESTION FOUR:

a. Fill in the correct response:

The odds that the blood samples taken from Bill Taylor's apartment belonged to Greg Johnson are one in ________.

a. 100,000
b. 50,000
c. 10,000
d. 5,000
e. 1,000
f. 500
g. 100
h. 50
i. 10
j. I cannot say

b. In your own words, explain what the phrase "blood type A2B occurs in .01% of the population of Boise, Idaho" implies about Greg Johnson:

____

____

____

____

____

____

____

QUESTION FIVE:

a. Based on all the evidence you heard in this trial, how much doubt is there in your mind that Greg is guilty of murdering Bill?

|____|____|____|____|____|____|____|____|____|____|
0 10 20 30 40 50 60 70 80 90 100

No doubt — Moderate doubt — Definite doubt

b. In your own words, explain what you understand by the phrase "beyond a reasonable doubt":

____

____

____

____

____

____

____

____

____

c. On the following scale, indicate how much doubt you would consider to be a "reasonable doubt", in accordance with instructions you received before determining your verdict.

0 10 20 30 40 50 60 70 80 90 100

No doubt	Moderate doubt	Definite doubt

(Please turn the page)

QUESTION SIX:

For questions a, b and c, please refer to the following guidelines in formulating your responses:

0 = totally inaccurate/unreliable
0 to 20 = very rarely accurate/reliable
20 to 40 = seldom accurate/reliable
40 to 60 = moderately accurate/reliable
60 to 80 = generally accurate/reliable
80 to 100 = most frequently accurate/reliable
100 = totally accurate/reliable

In your opinion:

a. how accurate or reliable is eyewitness testimony?

|____|____|____|____|____|____|____|____|____|____|
0 10 20 30 40 50 60 70 80 90 100

Inaccurate Accurate

b. how accurate are statistics?

|____|____|____|____|____|____|____|____|____|____|
0 10 20 30 40 50 60 70 80 90 100

Inaccurate Accurate

c. how accurate are tests of blood types?

|____|____|____|____|____|____|____|____|____|____|
0 10 20 30 40 50 60 70 80 90 100

Inaccurate Accurate

QUESTION SEVEN:

a. You have bet $100 that the blood sample taken from Bill's apartment hallway belongs to Greg Johnson. Your answers to (i) and (ii) must add up to 100.

(i) I have a _____% chance of winning back the $100 which I bet, plus another $100.

(ii) I have a _____% chance of losing the $100 which I bet.

TOTAL: 100%

b. Fill in the correct response:

Based on the evidence presented in the trial of Greg Johnson, approximately _________ people in Boise have blood type A2B?

a. 10
b. 50
c. 100
d. 500
e. 1,000
f. 5,000
g. 10,000
h. 50,000
i. 100,000
j. I cannot say how many

QUESTION EIGHT:

a. Select the statement which most accurately describes you:

In terms of understanding math, I ________ have difficulty.

a. never
b. rarely
c. sometimes
d. often
e. always

b. Select the statment which most accurately describes you:

In conversation, to support my arguments, I _________ use statistics.

a. never
b. rarely
c. sometimes
d. often
e. always

QUESTION NINE:

If the forensic experts had proven not only that the blood samples taken from Bill's apartment hallway matched Greg Johnson's bloodtype, but also that these samples could not possibly have come from any other person, based on the evidence presented in this trial, how much doubt would you have that Greg Johnson is guilty of murdering Bill Taylor?

|_____|_____|_____|_____|_____|_____|_____|_____|_____|_____|
0 10 20 30 40 50 60 70 80 90 100

No doubt — Moderate doubt — Definite doubt

QUESTION TEN:

a. Are you a registered voter? _________

b. Have you ever been called to serve as a juror? _________

c. What is your dominant language? _________________

d. Years of education since high school? _________

e. Most advanced educational qualification? _________

f. Age in years: _______ f. Gender: _______

APPENDIX B

Materials for Study Two

IN THE KING COUNTY SUPERIOR COURT
FOR THE STATE OF WASHINGTON

STATE OF WASHINGTON, Prosecutor, v. ARTHUR W. JACKSON, dba NW ALUMINUM CORP., Defendant.	SUMMARY OF TRIAL TRANSCRIPT

OPENING STATEMENTS (excerpts):

The Prosecution will try to prove that the defendant Arthur Jackson committed the crime of arson by knowingly and maliciously requesting or encouraging George Avery to cause a fire or explosion on his property, the NW Aluminum plant, with the intent to collect insurance proceeds in the sum of $1,667,000 from the American Insurance Company.

The Defendant, Arthur Jackson, will try to show that the fire which destroyed the Northwest Aluminum plant in Redmond on November 16, 1984 was accidental.

PROSECUTION'S CASE (excerpts):

Witness 1: MARIE WILLIAMS, former bookkeeper at NW Aluminum.

On direct examination, Marie Williams testifies as follows:

She worked for NW Aluminum for ten years. In 1983, company losses totaled about $500,000. Losses for the first nine months of 1984 amounted to $500,000. In July, 1984, Mr. Jackson told her that if the company did not pick up some new accounts before the end of the year, it would go under. He was very concerned about the company's financial problems.

In September, she overheard a conversation between Mr. Jackson and his new administrative aide, Sonia Peterson. Peterson offered to put Jackson in touch with someone named Avery, whom she said could solve his financial problems.

Bank loans to NW Aluminum in the sum of $400,000 were due on November 28, 1984. Williams and Jackson met with the bank manager to try and renegotiate the loans. The negotiations failed.

In the middle of October, 1984, Jackson met with George Avery in his office. After the meeting, Williams overheard Sonia Peterson describe Avery as "a torch."

On November 1, 1984, Williams signed a company check to George Avery in the amount of $5,000. Mr. Jackson said the payment was for plans for new aluminum tools and dies which Avery was designing for the manufacture of automobile parts. Up to that time, the company made siding for home construction.

On November 16, 1984, the day the aluminum plant burnt down, she worked late. Around 6:45 p.m. she heard the last work crew finish a rush job it was working on in the first floor machine shop. From her office window on the third floor, she saw several of the crew leaving the property. She was surprised to hear some machinery after that. On her way downstairs to check on this, she ran into Mr. Jackson on the stairwell. They had a brief conversation during which he told her she looked tired and suggested that she go home, which she did. She heard about the fire later that night on the 11:00 o'clock news.

<u>On cross examination</u>, Marie Williams testifies as follows:

She and Mr. Jackson had been lovers for four years. She broke off the relationship two years ago when she realized that Mr. Jackson was unwilling to leave his wife and marry her.

The day after the fire, she told Mr. Jackson she suspected he had paid Avery to burn the plant for the insurance money. Ten days later, he dismissed her from her job. After she was dismissed, she went to the police and told them all the facts she stated in court today.

She attended only one meeting with Mr. Anderson, the vice-president of the bank and Mr. Jackson during which loans were discussed. There may have been others.

A can of gasoline was usually kept in the first floor storage room for use in running the company lawnmower. (A diagram of the layout of the first floor of the Aluminum plant is attached as Exhibit A.) She regularly gave the gardener money from petty cash to purchase gasoline for this purpose. The last time the gardener purchased gasoline was in the first week of October, 1984. She does not know where he bought the gas.

She didn't know Avery personally, and had nothing to do with hiring of personnel. She smokes cigarettes and so did several of the workers at the plant.

Witness 2: <u>ROBERT ANDERSON</u>, Vice-President of Rainier Bank in Redmond:

<u>On direct examination</u>, Robert Anderson testifies as follows:

During October, 1984, Arthur Jackson sought to refinance a NW Aluminum loan in the amount of $400,000, due on November 28, 1984. He and Jackson met several times to discuss NW Aluminum finances. Anderson refused to renegotiate the loans. Jackson also wanted long-term financing for NW Aluminum to accomplish a transition to a new

product line.

On cross examination, Robert Anderson testifies as follows:

Mr. Jackson informed him of plans to manufacture aluminum parts for automobiles for companies such as General Motors and said he had hired a new designer for the project. Anderson requested copies of financial forecasts and designs of the new product line, but has never received these from NW Aluminum.

Witness 3: HOWARD OLSON, Chief Fire Marshall in Seattle.

On direct examination, Marshall Olson testifies as follows:

He has been employed as Chief Fire Marshall for 11 years. He conducted an investigation of the fire at the Redmond industrial plant on the evening of November 16, 1984. A passer-by called in a report of the blaze around 8:30 p.m. A few minutes later when four fire trucks arrived on the scene, they found a fire with large flames, rapidly spreading in a horizontal direction. It took five hours to extinguish the fire--an unusually long time.

The four story building was almost entirely destroyed. There was one fatality. The remains of a white Caucasian male later identified as George Avery were found beneath the debris of an explosion which occurred in the machine shop on the first floor. The cause of his death was blunt trauma, primarily to the chest, causing hemorrhage into the chest cavity and collapse of the lungs. Analyses of debris indicated that there was a "hot spot" in this area, probably the point of origin of the fire. Apparently, one of the machines in the shop was operating just prior to the fire. The fire hoses washed away a good deal of the evidence.

The rapid spread of the fire, and difficulty in extinguishing it point to the use of an accelerant. An examination of the concrete floor in the storeroom area showed "spalling" which often indicates that a flammable liquid was burned in this area. No container was found at the scene, but if the accelerant had been siphoned into a plastic container, the intensity of the fire would have burnt away all traces of a plastic container. Samples of gasoline found in the concrete floor were taken to the state laboratory for chemical analysis. Chromatographic tests revealed the presence of automotive gasoline in the concrete samples. Four different gasoline samples were analyzed to determine their similarity on the basis of twelve components: (See Exhibit B, attached.)

1. Gasoline from a can found in the first floor storage room.
2. Gasoline from the gas tank of Avery's car, standing in the parking lot outside the NW Aluminum plant.
3. Gasoline from a commercial pump within one mile of the site of the fire.
4. Gasoline from the concrete floor of the first floor storage room.

Test results indicated that the gasoline recovered from the concrete was dissimilar to that stored on the site in a can, possibly intended for lawnmower use. In other words, the gasoline in the can was not the source of the gas recovered from the concrete. Tests results also showed that the gasoline from a Chevron pump one mile from the plant was dissimilar to the gas recovered from the concrete floor.

The tests further revealed that there were no significant differences between the gas sampled from Avery's car and the samples recovered from the concrete floor. The odds against this result occurring by chance alone are 1 in 1000. Therefore, these two gases probably have a common origin.

He concluded that the fire was the result of arson because:

a) Avery was present in the building at the time of the fire;
b) The fire spread rapidly and was unusually well-ventilated;
c) Gasoline traces found in the concrete in the storeroom matched the gasoline found in Avery's automobile.

<u>On cross-examination, Marshall Olson</u> testifies that:

It is a standard procedure to investigate all large fires in the area in which there are any fatalities.

Investigation revealed that a small amount of gasoline was usually stored in the first floor storage room by the gardener to operate the company lawnmower. A gasoline can will explode in a fire, so long as it is not very full, because the vapor in the space between the liquid level and the top of the container expands when it is heated, resulting in a "vapor explosion". A vapor explosion can be powerful enough to blow out the windows of a building and ventilate the fire, making it burn more furiously. The debris of this fire made it impossible to say whether a vapor explosion did or did not occur in the early stages of the fire.

He cannot say for sure that the source of the gasoline samples found in the concrete floor was not someone who entered the premises after the fire was extinguished to deliberately create the impression of arson. A police officer was stationed at the site after the fire was extinguished. It is possible that there was there was a lapse in security. If so, an intruder might have splashed gasoline on the concrete floor before the samples were collected the next day for analysis by the state laboratory.

Many arsonists are trapped and killed in their own fires. Professional arsonists are unlikely to be caught in their own fires. Arsonists who die in their own fires are usually less skilled or less experienced.

Redirect examination of Marshall Olson:

The gasoline can found on the first floor was intact, and was in no way responsible for a vapor explosion. The scientific tests showed that gasoline in Avery's car matched the gasoline found in the concete floor. You cannot expect two samples of gasoline from a common source to look identical in every detail if one has been burned and the other not.

The prosecution rests.

DEFENDANT'S CASE:

Witness 1: ARTHUR JACKSON, President and sole shareholder of NW Aluminum.

On direct examination, Arthur Jackson testifies as follows:

He is 50 years old, currently separated from his wife. He holds a Master's Degree in Business Administration from the University of Washington. Before purchasing NW Aluminum 13 years ago, he worked as Assistant President for another aluminum company for three years.

NW Aluminum manufactured aluminum siding and windows for use in home construction. He employed about 40 people. Business was good until the housing market fell off significantly in 1982. The business had been losing money for the past two years. He personally had no financial difficulties because his wife had inherited $300,000 and he had personal assets of $50,000. His wife was not willing to loan funds to the business. NW Aluminum had bank loans of $400,000 due on November 28, 1984. He planned to negotiate new loans before that date. Ever since the plant was destroyed in the fire on November 16, NW Aluminum has not been open for business.

On the day of the fire, he went to the plant at about 7:00 p.m. to pick up some papers. He chatted briefly with Avery who was working late. George Avery was working on a special project to produce designs for a new line of products for the company. He hired Avery in October, on the recommendation of Sonia Peterson, his administrative aide. Avery was not on the regular company payroll, but was an independent consultant, retained at a fee of $5,000. Avery was a brilliant designer and had suggested using aluminum to replace steel in automobile parts. This could have saved the business. Jackson had every reason to believe the new product line would turn the company around. While the bank refused to renegotiate old loans, it was prepared to negotiate new loans on submission of Avery's new designs for retooling and marketing projections. Avery was busy preparing these designs on November 16.

Jackson bumped into Marie Williams on the stairs, just as he was leaving the plant. Then, around 10:00 p.m., the Redmond Fire Department called him at home and told him the plant was ablaze. He

went there immediately. He does not know how the fire started. Avery's death in the fire was an unspeakable tragedy. All the new designs and plans were also destroyed in the fire.

The insurance coverage on the plant was increased regularly every five years by about $500,000, to account for the effects of inflation. In October, 1984, he increased coverage from $1,125,000 to $1,667,000. NW Aluminum had been insured by the American Insurance Company for eleven years and paid all its premiums. The policy does not cover losses resulting from arson where the arson is caused by the deliberate acts of the insured or its agents. (A copy of the policy is attached as Exhibit C.)

On cross examination Arthur Jackson testifies as follows:

He dismissed Marie Williams for two reasons. First, she was becoming an embarrassment, and second, after the fire, business came to a halt, and he no longer needed a bookeeper. It is true that he had an affair with her which ended in 1983. Marie Williams was extremely jealous of Sonia Peterson, his administrative aide. Her jealousy was the reason he ended the relationship with her. After the fire, Marie told him she suspected him of arson. He didn't take this very seriously. He is trying to reconcile his relationship with his wife.

On January 1, 1979, he increased NW Aluminum's fire insurance from $835,000 to $1,125,000. The last increase was on October 1, 1984, six weeks before the fire.

Witness 2: SONIA PETERSON, Administrative Secretary to Arthur Jackson.

On direct examination Sonia Peterson testifies as follows:

She worked for Mr. Jackson for two years and is presently unemployed. She knew George Avery before when they both worked at another company, and she recommended him to Jackson as an experienced engineer who could design a new line of aluminum products for the company.

In November, Jackson was negotiating with the bank to secure a loan to cover new line of products.

She has never referred to George Avery as "a torch". Ms Williams is mistaken. Perhaps Ms. Williams overheard her telling Mr. Anderson that she had to see the Olympic torch when it was carried through Seattle before the Olympic games in 1984.

On cross examination Sonia Peterson testifies that:

Jackson never asked her to send any materials about the new designs to the bank manager. She was not at the plant evening of fire. She and Jackson were not having an affair. She believes Mr. Jackson is a good man, but he is quite a lot older than she is. She has a steady

boyfriend.

When Avery submitted his resume, she never verified his qualifications by calling the university from where he stated he received his engineering training to determine whether he was in fact a graduate engineer. She was familiar with his background and his reputation. Avery was a bit of a loner, and often worked as an independent contractor.

Witness 3: DAVID PINKUS, Fire Investigator

On direct examination, David Pinkus testifies that:

He worked for 20 years as deputy fire marshall in Seattle. Since then, he has been employed as a consulting fire investigator. He has studied modern methods of fire investigation, including chromatography.

He reviewed the reports compiled by Olson, the witness statements and also interviewed several employees at NW Aluminum. Based on this investigation, he concluded that the fire on November 16, 1984, must be listed as an accident.

The presence of gasoline at the scene of fire can be explained by fact that gardener regularly used gasoline to run the company lawnmower. This gasoline was stored in the first floor storeroom, close to where Avery's body was found. Finding one full can after the fire doesn't rule out the possibility that other cans containing less gasoline were also present in this area.

The time lapse between the extinguishment of the fire and the analysis of the gas samples was too great to permit a precise analysis. Even if scientific tests revealed that the samples of gasoline from Avery's car and those found in the fire debris were identical in every respect, this merely proves that both Avery and someone else purchased gasoline from a common supply source, and not necessarily that Avery was a hired arsonist.

Following the fire, there was sufficient time for someone to enter the premises and pour gasoline onto the concrete floor to make it look like arson. The scientific tests did not establish when the gasoline samples were placed at the site. The intruder could have taken the gasoline from Avery's car. Anyone filling up with gas at the same station as Avery at approximately the same time would have received gasoline virtually indistinguishable from his. Many other indistinguishable sources of gasoline are possible throughout the city of Redmond.

The unusually large flames and rapid fire were caused by the burning of the wooden building, not necessarily the splashing about of an accelerant. There is no question that the building was about 60 years old, and constructed primarily of wood. Furthermore, the floor plan was open, providing favorable ventilation conditions, leading to a

rapidly-spreading blaze.

It is highly unlikely that a professional arsonist would get caught in his own fire. The death of Avery in the fire points to an accidental explosion which probably started the fire. Evidence of accidental causes was washed away by hoses. Any of the workers in building could have left a cigarette butt smouldering; machinery could have had faulty wiring and sparked a fire; it is impossible to say. Until accidental causes are ruled out, no finding of arson can be made. While arson is possible, it is not probable. He believes that the fire was accidental because:

a) Presence of Avery's body at the scene of the fire.
b) Presence in building shortly before start of fire of numerous employees who could have left machinery running, left appliances on, left cigarette butts, etc.
c) Age of wooden building and open, ventilated floor-plan.
d) Presence in building of flammable liquids such as gasoline.

<u>On cross-examination</u> David Pinkus testifies that:

He never visited the scene of the fire to inspect the site in person.

The term "torch" is a slang expression to describe an arsonist, or someone who deliberately starts a fire. A common motive for arson is recovery of insurance proceeds.

He respects the work of Marshall Olson, but disagrees with his conclusion. Chief Olson is in the business of finding arsonists. He, Pinkus, is an impartial investigator.

<u>The defense rests.</u>

<u>PROSECUTION'S CLOSING ARGUMENT</u>:

The evidence shows that Arthur Jackson committed arson. He had ample motive for the crime. The company had very serious financial problems and no way out of them. On top of two years of drastic losses, it owed $400,000 in two weeks - an amount it could not pay. The bank had refused to renegotiate the loans. Money was also needed for the future operating expenses. Robert Anderson had asked for evidence of new product designs and financial forecasts, and was never provided with either. Mrs. Jackson certainly wasn't going to help the company. She was separated from Arthur Jackson. The insurance money was just enough to bail NW Aluminum out of trouble. This fact is not coincidental, for Jackson himself increased the amount of the insurance to meet those needs just 6 weeks prior to the fire. Usually, he would increase insurance on January 1st each year. Why did he suddenly increase the insurance in September and not January?

There is no reliable evidence that Avery was a qualified designer, nor that he produced any plans or designs. Peterson conveniently omitted checking on Avery's credentials and references, probably because she knew he was an arsonist. The testimony of Marie Williams that she overheard them talking about the "torch" supports this. Consider also the highly irregular way in which Avery was paid - check for $5,000 from Jackson two weeks before the fire.

Based on scientific tests establishing a crucial link between Avery and the cause of the blaze - the matching gasoline samples, Chief Olson was satisfied that the fire was not accidental. This was clearly arson.

DEFENDANT'S CLOSING ARGUMENT:

There is not sufficient evidence of arson to deny Arthur Jackson his insurance proceeds. You can't believe Marie Williams because she was motivated by jealousy and anger to ruin Arthur Jackson after he failed to leave his wife and marry her. Who is to say whether she did not seize this opportunity to splash gasoline at the scene of the fire just to make it look like arson, and ruin his life forever. Someone may have wanted to set up Avery. Alternate motives of other characters have not been adequately pursued and eliminated.

The evidence fell far short of establishing Avery as an arsonist. At most, Olson's tests established only that gasoline in Avery's car and gasoline found on the concrete floor of the storeroom was purchased from the same source. This does not indicate arson. Thre was every reason to believe fire was accidental: the building was extremely old, the layout promoted fires, the wiring was old, employees worked late and left in a hurry, perhaps without turning off the machines, or properly extinguishing a lst cigarette. Even Marie Williams admitted hearing machinery running after the crew left. She too was also a cigarette smoker. Until accident is ruled out, there is no arson. Chief Olson was looking for signs of arson. Pinkus was looking impartially at the evidence. Even Olson agreed that arsonists rarely die in their own fires.

Many engineers work as independent contractors and are paid a retainer in advance. Even Anderson knew Avery was hired to work on designs. Jackson was a seasoned and resourceful employer. The clock had not run out on the available time to negotiate new loans, to bring designs to the bank and secure financing for the future. Why would he destroy what he spent his life buiding? This was an accidental fire, and the insurance company should pay up.

INSTRUCTIONS FROM THE JUDGE TO THE JURY:

The parties to this case are the State of Washington, the prosecutor, and Arthur Jackson, the defendant. The aluminum fabrication plant owned by defendant and insured by the American Insurance Company was destroyed by fire on November 16, 1984. The insurance policy contains a clause which provides that the company will not be liable for any loss caused by or resulting from arson if that arson is the result of acts of the defendant or its agents. The defendant claims the fire was an accident and that he is entitled to recover $1,667,000 from the company. The state has prosecuted Mr. Jackson for arson.

Neither opening statements nor closing arguments are evidence, and any statement or argument made by the attorneys that is not based on the evidence should be disregarded by you in reaching a verdict.

You are the sole judges of the credibility of the witnesses and of the weight to be given to the testimony of each witness. You may take into account the ability and opportunity of the witness to observe, and any interest, bias, or prejudice he or she may have; the reasonableness of the testimony considered in the light of all the evidence, and any other factors that bear on the believability and weight of the witness's testimony.

The law allows experts to express an opinion on subjects involving their special knowledge, training and skill, experience or research. While their opinions are allowed to be given, it is entirely within the province of the jury to determine what weight shall be given to their testimony. Their testimony is to be weighed as that of any other witness.

The law recognizes two kinds of evidence: direct and circumstantial, and each should be considered according to whatever weight or value it may have. Direct evidence if true, establishes a fact itself. Circumstantial evidence is proof of facts or circumstances which give rise to a reasonable inference of other facts. Circumstantial evidence proves a fact indirectly in that it follows from other facts or circumstances according to common experience and observations in life.

The burden of proof is on the state to prove the defendant guilty beyond a reasonable doubt. For the state to prevail, you must find that both of the following propositions were proven:

a. Jackson or his agents knowingly and maliciously caused a fire or explosion on the NW Aluminum property
b. with intent to collect insurance proceeds.

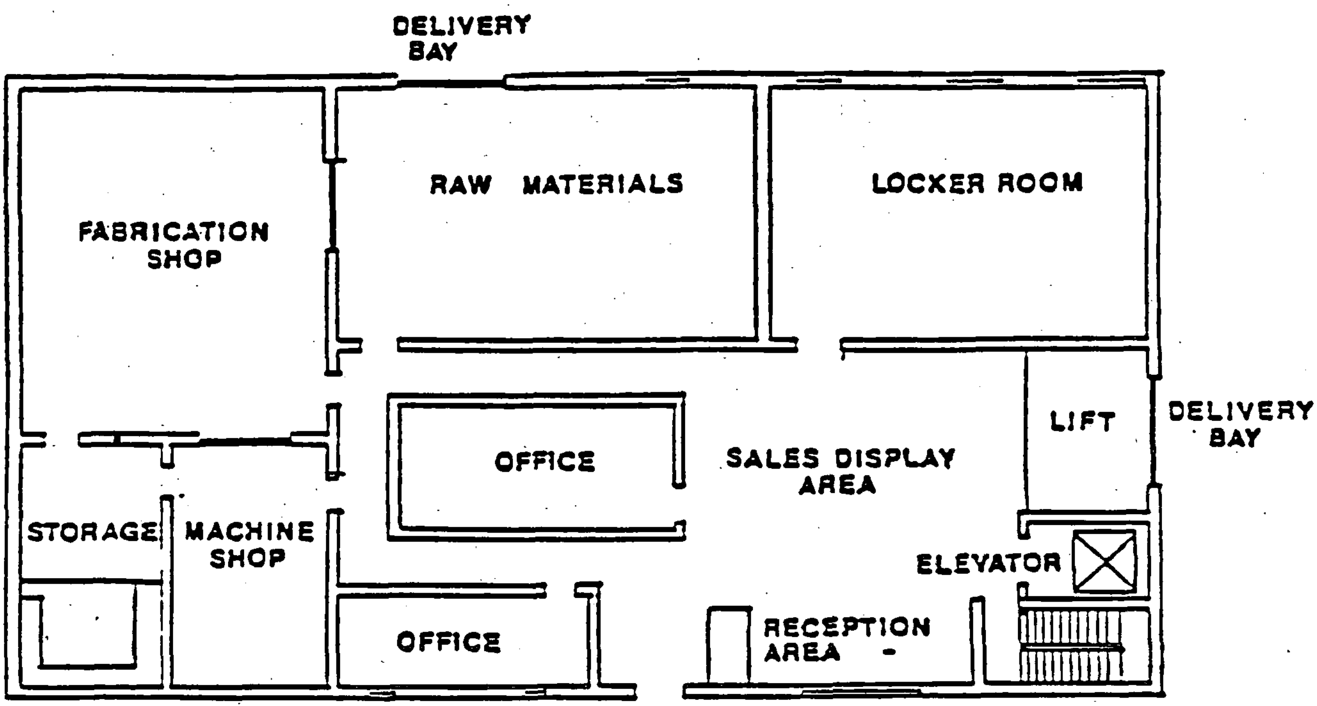
The defendant has entered a plea of not guilty. The defendant is presumed innocent and is not required to prove his innocence. Reasonable doubt means a doubt based upon reason and common sense that arises from a fair and rational consideration of all the evidence or

lack of evidence in the case. It is a doubt that is not vague, speculative or imaginary, but a doubt as would cause a reasonable person to hesitate to act in matters of importance to themselves.

If you find that either of these propositions has not been proven beyond a reasonable doubt, then your verdict must be for the defendant, i.e., not guilty. If you find that both of these propositions have been proven beyond a reasonable doubt, your verdict must be for the state, i.e., guilty.

EXHIBIT A:

FIRST FLOOR LAYOUT
NW ALUMINUM PLANT
Redmond, Washington



FIRST FLOOR LAYOUT

EXHIBIT B:

CHROMATOGRAPHIC TESTS ON GASOLINE SAMPLES

Sample	Source
1 -	Can of gasoline found in storage room at NW Aluminum
2 -	Gas tank of George Avery's car in NW Aluminum parking lot
3 -	Chevron gas station one mile from NW Aluminum plant
4 -	Concrete floor of NW Aluminum storage room

Results of chromatographic analysis on the basis of twelve components:

Samples 1 and 2:	dissimilar
Samples 1 and 3:	dissimilar
Samples 1 and 4:	dissimilar
Samples 2 and 3:	dissimilar
Samples 2 and 4:	no significant difference
Samples 3 and 4:	dissimilar

The likelihood of samples 2 and 4 matching by chance alone: .001%

CASE STUDY: COMPARISON OF GASOLINES IN NW ALUMINUM PLANT FIRE

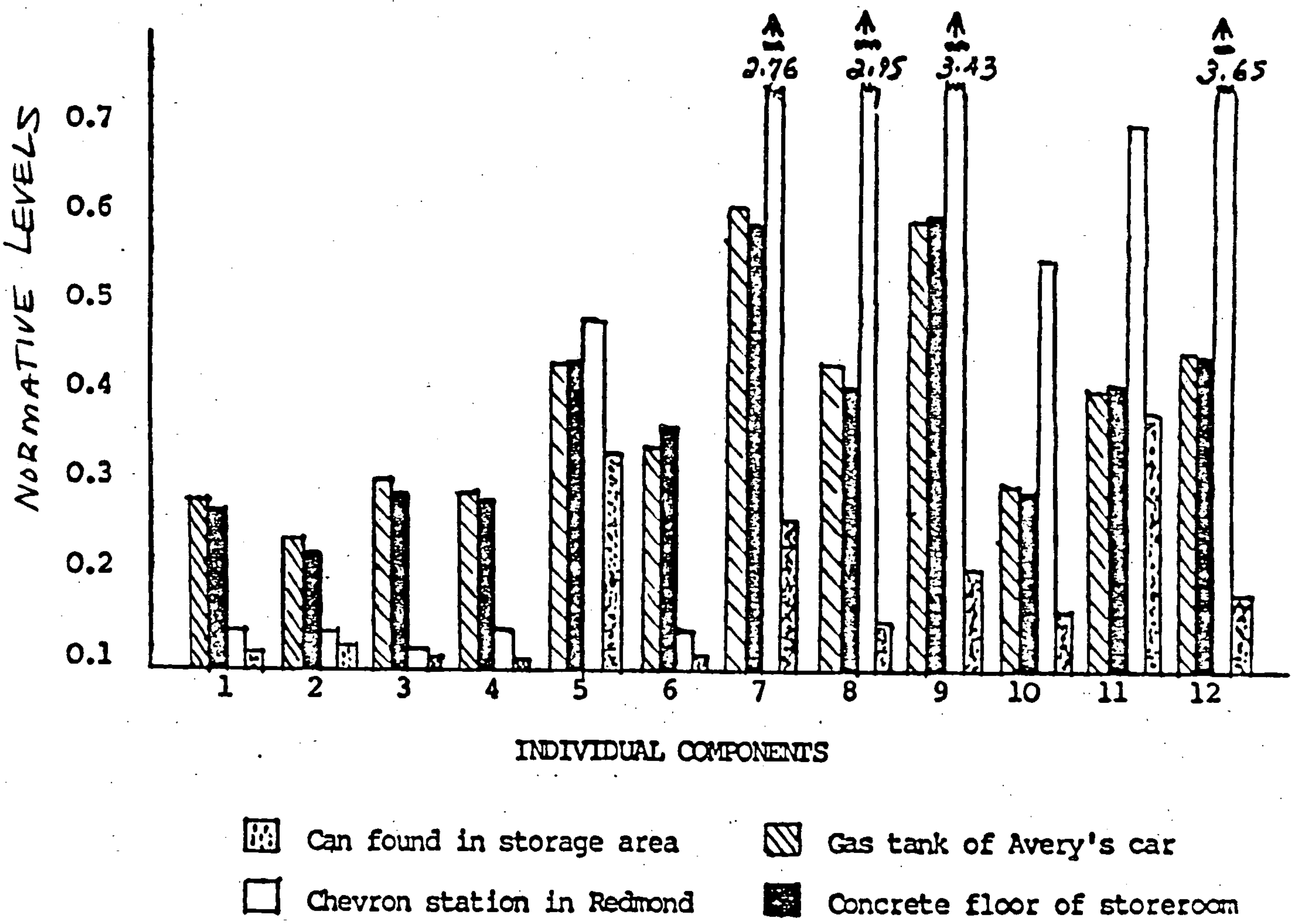


EXHIBIT C:

FIRE INSURANCE POLICY

The American Insurance Company

Policy No: 951946

AGREEMENT between the American Fire Insurance Company (hereinafter the "Company") and Northwest Aluminum Corporation, (hereinafter the "insured"):

This policy is to take effect on January 1, 1973.

FACE AMOUNT: $835,000

INSURED PREMISES: The plant and property of the insured located at 19707 River Road, Redmond, Washington.

ENDORSEMENTS

FACE AMOUNT: Increased to $1,125,000
January 1, 1979.

FACE AMOUNT: Increased to $1,667,000
October 1, 1984.

[Clauses 1 to 8 of the standard fire insurance policy are omitted.]

Clause 9 of the policy reads as follows:

9. ARSON: The company shall not be liable for loss caused by or resulting from arson where the same was occasioned by the deliberate acts of the insured or any agents thereof.

VERDICT

Do you find beyond a reasonable doubt that Jackson or his agents knowingly and maliciously caused a fire or explosion on the NW Aluminum property on November 16, 1984? Yes No

If yes, do you find beyond a reasonable doubt that this was done with the intent to collect insurance proceeds from the American Insurance Company in the amount of $1,667,000? Yes No

Do you find Arthur Jackson, dba Northwest Aluminum Corporation, guilty of arson? Yes No

Using the scale below, indicate how confident you are that your verdict is correct? (Write percentage in the response blank.)

Ans: ____

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|

0 10 20 30 40 50 60 70 80 90 100

Not at all confident	Not very confident	Somewhat confident	Quite confident	Extremely confident

What three specific facts presented in the trial were most important to your verdict? (List the most important first, then the second most important, etc.)

1. ______________________________
2. ______________________________
3. ______________________________

Summarize the reasons for your verdict in this case, and note any particular observations you may have about the case.

____

____

____

____

____

The most important instruction from the judge was to:

____

The most important exhibit in this trial was:

____

An example of direct evidence in this trial was:

____

An example of circumstantial evidence in this trial was:

____

Suppose you have 100 points to divide among the 6 witnesses who testified. Indicate the relative weight you gave to the testimony of each in reaching your verdict. The total must add up to 100.

Witness	Weight
Marie Williams	_____
Robert Anderson	_____
Howard Olson	_____
Arthur Jackson	_____
Sonia Peterson	_____
David Pinkus	_____
Total:	100

Here are some questions about the evidence presented during the trial. Please fill in an answer to every question, even if you must guess:

NW Aluminum Corp. was in financial trouble to the extent of $________

Loan payments were due ________ weeks after the fire.

Usually, Jackson increased NW Aluminum insurance every ___________

The total value of the insurance policy was $ ___________.

The gasoline samples were analyzed by means of __________________ tests.

The gasoline samples were compared on ____________________ components.

The number of gasoline samples compared was ______________________ .

The odds of two gasoline samples matching by chance alone depend on

____

How similar were gasoline samples from Avery's car and the storeroom floor?

____

____

How similar were gasoline samples from the Chevron station and the storeroom floor?

____

____

There is a _____ per cent chance of matching gas samples by chance alone.

To ascertain the likelihood of matching gas samples by chance alone, you need to sample at least

____

The odds against matching gas samples by chance alone are

____

Using the following scale, indicate how believable each of the following is: (Fill the percentage in the blank.)

0	10	20	30	40	50	60	70	80	90	100
Extremely unbelievable				Uncertain						Extremely believable

Testimony of Marie Williams ______

Testimony of Arthur Jackson ______

Testimony of Robert Anderson ______

Testimony of Sonia Peterson ______

Testimony of Howard Olson ______

Testimony of David Pinkus ______

Using the following scale, indicate the likelihood of each of the following: (Fill the percentage in the blank at the right of the page.)

0	10	20	30	40	50	60	70	80	90	100
Extremely unlikely				More likely than not						Extremely likely

Jackson planned to renegotiate NW Aluminum's bank loans before November 28.

The rapid blaze was attributable to the use of an accelerant such as automotive gasoline.

Williams knew more about the causes of the fire than she stated at trial.

The similarity of gasoline samples from the storeroom floor and from Avery's gas tank indicated arson.

Jackson wanted the insurance money to pay off the company debts.

The fire on November 16 was deliberately caused.

Avery was an experienced tool and die designer.

Here's the scale again so you don't have to page back to see it:

```
|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
0   10    20    30    40    50    60    70    80    90   100
Extremely              More likely                  Extremely
unlikely               than not                     likely
```

Jackson ended the relationship with Williams.

After the fire, an intruder siphoned gasoline from Avery's car.

Jackson planned to submit market studies, sales projections and designs of the new product line to Mr. Anderson to obtain long-term financing for the transition to the new product line at NW Aluminum.

Sonia Peterson described Avery as a "torch".

Gasoline samples from the storeroom floor and from Avery's gas tank had a common source.

Avery was an experienced arsonist.

Professional arsonists are caught in their own fires.

Someone entered NW Aluminum premises after the fire was extinguished and poured automotive gasoline on the concrete floor.

The rapid blaze was attributable to the open plan and wooden frame of the building.

Chance factors account for the similarity of gasoline samples from the storeroom floor and from Avery's gas tank.

A vapor explosion occurred in the early stages of the fire.

Peterson knew more about the causes of the fire than she stated at trial.

Avery and another unidentified potential suspect purchased gasoline from the same source.

The fire on November 16 was accidentally caused.

Police security lapsed after the fire was extinguished.

Avery was an inexperienced arsonist.

Here's the scale again so you don't have to page back to see it:

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
0 10 20 30 40 50 60 70 80 90 100

Extremely unlikely	More likely than not	Extremely likely

Williams ended relationship with Jackson.

Peterson discussed the Olympic torch parade with Jackson.

Marie Williams was determined to ruin Arthur Jackson.

Sonia Peterson and Arthur Jackson were lovers.

Suppose you bet $100 that Avery siphoned gasoline from the tank of his car and poured it on the storeroom floor of the the NW Aluminum plant. Your answers to (i) and (ii) below must add up to 100.

(i) You have a _____% chance of winning back the $100 which you bet, plus another $100.

(ii) You have a _____% chance of losing the $100 which you bet.

TOTAL: 100%

Using the following scale indicate how reliable each of the following is: (Fill the percentage in the blank.)

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|
0 10 20 30 40 50 60 70 80 90 100

Extremely unreliable	Uncertain	Extremely reliable

Testimony of laywitnesses _____

Chromatography tests _____

Direct evidence _____

Circumstantial evidence _____

Gasoline frequency studies _____

Testimony of experts _____

For each question, circle the answer that best describes your opinion or feeling:

SD = strongly disagree SA = strongly agree N = neutral
D = disagree A = agree

I have had little use for mathematics since I left school.	SD D N A SA
Mathematics is enjoyable and stimulating to me.	SD D N A SA
I don't think I could do advanced mathematics.	SD D N A SA
When a math question is left unanswered, I continue to think about it afterward.	SD D N A SA
Mathematics is not important to me in my life's work.	SD D N A SA
It doesn't bother me at all to solve math problems.	SD D N A SA
I almost never have been at ease during math tests.	SD D N A SA
I am challenged by math problems I can't understand immediately.	SD D N A SA
Mathematics usually makes me feel uncomfortable and nervous.	SD D N A SA
I'm no good in math.	SD D N A SA
I need a firm mastery of mathematics for my present work.	SD D N A SA
I get a sinking feeling when I think of trying to solve math problems.	SD D N A SA
A math test would scare me.	SD D N A SA
I am sure I can do advanced work in mathematics.	SD D N A SA
Mathematics usually makes me feel uneasy and confused.	SD D N A SA
I like math puzzles.	SD D N A SA
When a math problem arises that I can't immediately solve, I stick with it until I have the solution.	SD D N A SA
I use mathematics in many ways.	SD D N A SA
I don't understand how some people can spend so much time on math, and seem to enjoy it.	SD D N A SA
I have a lot of self confidence when it comes to math.	SD D N A SA

My mind goes blank and I am unable to think clearly when working mathematics.	SD D N A SA
I studied mathematics because I know how useful it is.	SD D N A SA
I do as little math as possible.	SD D N A SA
I would rather have someone give me a solution to a difficult math problem than to have to work it out myself.	SD D N A SA
For some reason, even though I try, math seems unusually hard for me.	SD D N A SA
Knowing mathematics helps me earn a living.	SD D N A SA
I'm not the type to do well in statistics.	SD D N A SA
Mathematics is a worthwhile and useful subject.	SD D N A SA
I see mathematics as a subject I rarely use in my daily life.	SD D N A SA
Generally I have felt secure about attempting mathematics.	SD D N A SA
Studying mathematics is a waste of time.	SD D N A SA
In terms of my adult life, it was important for me to do well in mathematics in high school.	SD D N A SA
I think I can handle more difficult mathematics.	SD D N A SA
I have almost never been distressed during a math test.	SD D N A SA
Mathematics makes me feel uncomfortable, restless, irritable and impatient.	SD D N A SA
I need mathematics for my present work.	SD D N A SA
Most subjects I can handle OK, but I have a knack for flubbing up math.	SD D N A SA
Once I start trying to work on a math puzzle, I find it hard to stop.	SD D N A SA
Math has been one of my worst subjects.	SD D N A SA
I haven't usually worried about being able to solve math problems.	SD D N A SA
Math puzzles are boring.	SD D N A SA
I do well in mathematics.	SD D N A SA
I usually have been at ease during math classes.	SD D N A SA

Figuring out mathematical problems does not appeal to me.	SD D N A SA
I am sure I can learn statistics.	SD D N A SA
Mathematics is of no relevance to my life.	SD D N A SA
Math doesn't scare me at all.	SD D N A SA

Please provide the following personal information: (Circle the correct and answer or fill in the blank.)

Have you ever served on a jury? Yes No

What is your home language? ______________________

Did you complete high school? Yes No

Have you studied mathematics since high school? Yes No

Have you ever studied statistics? Yes No

What is your most advanced educational qualification?

____

What is the most advanced mathematics course taken:

____

What is the most advanced statistics course taken:

____

Gender? Male Female

Year born? __________

High-tech process helps convict man of arson

by Julie Emery
Times staff reporter

For the first time in the state, a high-tech process known as capillary gas chromatography has been used in a court to help convict a man charged with arson.

Jurors put their faith in the process Tuesday when they found a Bellevue man, Edward J. Michel, 48, guilty of first-degree arson for torching the garage of a family home.

Prosecutors and police say the gas chromatography process helps them link gasoline to arsons much in the same way a bullet can be linked to a gun used in a murder, said Senior Deputy Prosecutor William Downing.

In Michel's case, staffers at the Western Washington State Crime Laboratory in Seattle used the technology to physically link gasoline Michel placed on newspapers in the family's garage to gasoline found in the trunk of his car. They were able to eliminate gasoline in cars in the garage as a possible cause of the blaze.

The fire caused more than $100,000 damage to the family's home and $25,000 damage to cars parked in the garage, Downing said.

Dale Mann of the crime lab said staffers refined existing chromatography techniques used in medicine, environmental work and the petroleum industries to make the system court-proof.

Gas chromatography involves separating a mixture of individual chemicals and generating a specific pattern on graph paper that represents the mixture. The pattern that emerges for kerosene, for example, is different than the pattern set for gasoline.

The lab was able to say that the automotive gas that caused the garage fire probably was the same as that found in Michel's trunk. The inability to differentiate between Michel's trunk gas and the fire gas suggested the two gasolines stemmed from a common source, Mann said.

The sample of gasoline used in a fire has to fit certain rigid criteria before it can be compared. For instance, if the fire burns too long, not enough gasoline residue remains to make a comparison.

"Also, if the gasoline is poured on a substance that creates a large chemical background, then that will obscure the details that we need to make the comparison," Mann said. "Those are the two primary things that work against us."

Conditions clicked in the Michel case. Bellevue police and firefighters got to the scene and to the suspect quickly, Downing said. Thus residue at the fire scene could be matched with the gasoline in Michel's possession.

Each juror was given a set of charts, known as chromatograms, and followed along as Mann testified on the crime lab's findings.

Chromotography has been used by the lab in arsons before to identify the nature of a fuel used in a fire, but this is the first time the staffers have been able to testify in court to match a known source. The lab's research has been under way 2½ years.

In the past, police and prosecutors have had to rely largely on eyewitness accounts or motives to convince jurors to find accused arsonists guilty.

Seattle Times, April 13, 1984.

APPENDIX D: SAMPLE CONSENT FORM

Elizabeth Loftus, Professor of Psychology, 543-2640
Jane Goodman, Ph.D. Candidate, 545-2973

This research is being conducted by the University of Washington in conjunction with the Superior Court Judges and Administrators. The purpose of this study is to aid the courts in making changes in certain trial procedures. We will explain the study in more detail and answer any questions you may have when it is completed.

Participating in the study will involve reading a brief trial summary and completing a questionnaire about the summary. Your participation is strictly voluntary, and your answers will remain anonymous. You will not miss a chance to be called for jury selection. If your name is called for a trial, you will be notified here. The study will last about one and a half (1 1/2) hours. Only the researchers listed above will have access to the data and the data will be retained for approximately 9 months.

Your participation is greatly appreciated. Thank you.

____
Jane Goodman

____
Date

I agree to participate in this study with the understanding that I may withdraw at any time without penalty. I have had an opportunity to ask questions and understand that future questions I may have about the research or subjects' rights will be answered by one of the investigators listed above.

____
Signature

____
Date

Copies to: Subject
Investigator's file

VITA

Jane Goodman was born to Roy and Myrtle Fraser on February 17, 1952 in Johannesburg, South Africa. She attended Parktown High School for Girls in Parkview, Johannesburg, then completed a B.A. Degree and T.T.H.D. (Master's Degree) at the University of the Witwatersrand and the Johannesburg College of Education. She taught English literature, English as a second language, and French language and literature in Johannesburg before immigrating to the United States of America in 1976. She settled in Seattle, Washington, and in 1980, began studying law and psychology.

Curriculum Vita JANE GOODMAN

PERSONAL DATA:

Campus Address: Dept. of Psychology, NI-25
Univ. of Washington
Seattle, Washington 98125
(206) 545-2973

Law Office: U.S. Equal Employment Opportunity Commission
Seventh Floor, 1321 Second Avenue
Seattle, Washington 98101
(206) 442-4855

Social Security Number: 536-78-0448

EDUCATION:

1986 Ph.D. University of Washington.

1983 J.D., cum laude, Univ. of Puget Sound Law School. Admitted to Washington State Bar and Federal Dist. Court, Western Dist. of Washington, 1983.

1973 T.T.H.D., with distinction, Witwatersrand University & Transvaal Education Department, Rep. of S. Africa.

1972 B.A., Witwatersrand University, English & French,

HONORS AND AWARDS

U.S. Equal Employment Opportunity Commiss'n Special Commendation, 1986.
National Institute of Justice, Research Fellowship, 1985-1986.
American Psychological Association, Div. 41, Grant-in-Aid, 1984.
American Jurisprudence Prize for Criminal Law, Investigative, 1983.
U. of Washington Graduate School Ph.D. Scholarship, 1982-1983.
American Jurisprudence Prize for Constitutional Law, 1982.
Conrad Linder Memorial Award for English Writing, 1973.
Sir John Adamson Prize for Academic Achievements, 1973.
Transvaal Education Department Merit Awards, 1971, 1972.
Johannesburg Junior City Council Alderman, 1969-1971.
Johannesburg Youth Council International Exchange Student, 1969.
Johannesburg Junior City Council Elected Representative 1968-1969.

PUBLICATIONS:

Goodman, J., & Andrews, N. B. (1986, May). Negotiation Strategies in the Workplace. Paper presented at the Puget Sound Chapter of the Association of Women in Computing, Tukwila, WA.

Greene, E., Goodman, J, & Loftus, E. F. (1986, March). Juror competency in complex cases. Paper presented at the annual meeting of the Law and Psychology Association, Division 41 of the American Psychological Association, Tucson, Arizona.

Goodman, J., & Nickerson, P. H. (1985, December). Book Review: Social Science in Law: Cases and Materials, by J. Monahan & L. Walker. Trial, pp. 85-88.

Goodman, J., Greene, E. L., & Loftus, E. F. (1985, November). Juror comprehension of complex products cases. Trial, pp. 65-74.

Andrews, N. B., & Goodman, J. (1985). Negotiations tactics. Proceedings of the Annual Meeting of the Association of Women in Computing, Chicago, Illinois. (In press.)

Goodman, J. (1985, June). Scientific evidence: Jurors' inferential problems. Paper presented at the Annual Meeting of the Law and Society Association, San Diego, California.

Goodman, J. (1985, June). Linguistic universals in legal language simplification. Paper presented at the Annual Meeting of the Law and Society Association, San Diego, California.

Goodman, J., & Greene, E. L. (1984). Book Review: Making jury instructions understandable by Amiram Elwork, Bruce Sales & James J. Alfini. Florida Bar Journal, 11, 710-711.

Loftus, E. F., & Goodman, J. (1984). Techniques for direct and cross-examination of witnesses. In S. Kassin & L. Wrightsman, (Eds.) The psychology of evidence and courtroom procedure. Beverly Hills, CA: Sage.

Goodman, J., & Loftus, E. F. (1984, April). Social science looks at witness examination. Trial (pp. 52-57).

Loftus, E. F., Goodman, J., & Nagatkin, C. (1983). Examining witnesses - Good Advice and Bad. In R. J. Matlon & R. J. Crawford (Eds.), Communication strategies in the practice of lawyering, Annandale, VA: Speech Communication Association.

Goodman, J. (1983, February). Constitutional Protection for Language Rights and Linguistic Minorities. Paper presented at the annual meeting of the National Association for Bilingual Education, Washington D.C.

Editor, (1986-1987) Washington Lawyers Practice Manual, (Vol. 1-V), Seattle-King County Bar Association, Seattle, Washington.

RESEARCH EXPERIENCE:

Aug. 1986 -present	Research Associate, Psychology Dept., Univ. of Washington. Comprehension of Statistical Assessments in Civil Litigation. (E. F. Loftus.)
Sep. 1984 present	Research Associate, American Bar Association Litigation -- Section Jury Comprehension Committee, Chicago, Illinois. Juror competency in complex cases. (D. Margolis.)
Fall 1984- Fall 1985	Researcher and Writer for Integrated Perceptual Designers Project sponsored by the U.S. Air Force. (J. Lincoln.)
Feb. 1984- May, 1984	Research Assistant, Graham & Dunn, Seattle, Washington. Civil damage awards, real property. (M. Kipling.)
Jan. 1983- Jul. 1984	Legal Intern & Associate, Bricklin & Gendler, Seattle, Wa. Civil rights, family law, environmental law. (M. Gendler.)
May 1982- Aug. 1982	Legal Intern, Law Clinic, U. of Puget Sound, Tacoma, Wa. Criminal misdemeanors, family law, disabilities law.
Sep. 1981- Jun. 1982	Research Assistant, U. of Puget Sound Law School, Tacoma. Update book, <u>Jury selection in criminal and civil trials</u>, (1984). (A. F. Ginger.)

TEACHING EXPERIENCE:

Spr. 1985	Workshop Instructor, City of Seattle, Seattle, Writing for Managers and Supervisors.
Jan. 1974- Nov. 1975	Director of French Department, Sandown High School, Sandton, Transvaal Ed. Dept., Rep. of S. Africa.
Win. 1972- Win. 1973	English and French Language Arts Instructor, Johannesburg, Transvaal Ed. Dept., Rep. of S. Africa.
1970-1973	Instructor of E.S.L., English literature, Cresthill School of Languages, Johannesburg, Rep. of S. Africa.

PROFESSIONAL ASSOCIATIONS AND MEMBERSHIPS:

American Bar Association
American Judicature Society
American Psychological Association, Div. 9, 21, 23 and 41
American Trial Lawyers Association
Association of Women in Computing, Puget Sound Chapter
Law and Society Association
National Association for Bilingual Education
Seattle-King County Bar Association
Committee to Study Feasibility of Programs for Black S. Africans
Washington State Bar Association