Interpretation of statistical evidence in criminal trials - The Prosecutor's Fallacy and the Defense Attorney's Fallacy

Article *in* Law and Human Behavior · September 1987

DOI: 10.1007/BF01044641

William C. Thompson
University of California, Irvine

Law and Human Behavior, Vol. 11, No. 3, 1987

Interpretation of Statistical Evidence in Criminal Trials

The Prosecutor's Fallacy and the Defense Attorney's Fallacy*

William C. Thompson† and Edward L. Schumann†

In criminal cases where the evidence shows a match between the defendant and the perpetrator on some characteristic, the jury often receives statistical evidence on the incidence rate of the "matching" characteristic. Two experiments tested undergraduates' ability to use such evidence appropriately when judging the probable guilt of a criminal suspect based on written descriptions of evidence. Experiment 1 varied whether incidence rate statistics were presented as conditional probabilities or as percentages, and found the former promoted inferential errors favoring the prosecution while the latter produced more errors favoring the defense. Experiment 2 exposed subjects to two fallacious arguments on how to interpret the statistical evidence. The majority of subjects failed to detect the error in one or both of the arguments and made judgments consistent with fallacious reasoning. In both experiments a comparison of subjects' judgments to Bayesian norms revealed a general tendency to underutilize the statistical evidence. Theoretical and legal implications of these results are discussed.

INTRODUCTION

Crime laboratories often play an important role in the identification of criminal suspects (Saferstein, 1977; Schroeder, 1977; Giannelli, 1983). Laboratory tests

* This research was supported by grants to the first author from the National Science Foundation (No. SES 86-05323) and the UCI Academic Senate Committee on Research. The authors wish to thank Karen Rook and Robyn M. Dawes for comments on earlier versions of this article and John VanVlear for help in collecting data.

† Program in Social Ecology, University of California, Irvine. Correspondence should be sent to William C. Thompson, Program in Social Ecology, University of California, Irvine, California 92717.

may show, for example, that blood shed by the perpetrator at the scene of the crime matches the suspect's blood type (Jonakait, 1983), that a hair pulled from the head of the perpetrator matches samples of the suspect's hair (Note, 1983), or that carpet fibers found on the victim's body match the carpet in the suspect's apartment (Imwinkelreid, 1982b). Testimony about "matches" found through these comparisons is called associative evidence (Stoney, 1984). It is increasingly common in criminal trials where the defendant's identity is at issue (Imwinkelreid, 1982b; Schroeder, 1977; Peterson et al., 1984).

Associative evidence is sometimes accompanied by statistical testimony about the *incidence rate* of the "matching" characteristic. Where tests show the defendant and perpetrator share the same blood type, for example, an expert may provide information on the percentage of people in the general population who possess that blood type (e.g., Grunbaum, Selvin, Pace, & Black, 1978). Where microscopic comparisons reveal a match between the defendant's hair and samples of the perpetrator's hair, the expert may provide information on the incidence rate of such "matches" among hairs drawn from different individuals (Gaudette & Keeping, 1974). During the past 15 years, forensic scientists have devoted much effort to studying the incidence rate of various characteristics of hair (Gaudette & Keeping, 1974), soil (Saferstein, 1977, pp. 63–64), glass (Fong, 1973; Davis & DeHaan, 1977), paint (Pearson et al., 1971), and bodily fluids (Owens & Smalldon, 1975; Briggs, 1978; Gettinby, 1984). Statistical data from this literature are increasingly presented in criminal trials (Imwinkelried, 1982b; Note, 1983). One legal commentator, discussing research on blood typing, concluded that "our criminal justice system is now at the threshold of an explosion in the presentation of mathematical testimony" (Jonakait, 1983, p. 369).

The reaction of appellate courts to this type of evidence has been divided. The conflict stems largely from differing assumptions about the way jurors respond to incidence rate statistics. A few appellate courts have rejected such evidence on the grounds that jurors are likely to greatly overestimate its value (*People v. Robinson*, 1970; *People v. Macedono*, 1977; *State v. Carlson*, 1978; *People v. McMillen*, 1984). The majority of jurisdictions, however, admit such evidence on the grounds that jurors are unlikely to find it misleading and will give it appropriate weight (Annotation, 1980; Jonakait, 1983). Legal commentary on the issue appears divided between those who argue that statistical evidence may have an exaggerated impact on the jury (Tribe, 1971), and those who argue that statistical evidence is likely to be underutilized (Finkelstein & Fairley, 1970; Saks & Kidd, 1980).

Which of these positions is correct? Although no studies have examined people's evaluation of incidence rate statistics directly, research does exist on people's reactions to similar types of statistical information. A number of studies have shown, for example, that when people are asked to judge the likelihood of an event they often ignore or underutilize statistics on the base rate frequency of that event (for reviews see Bar-Hillel, 1980; Borgida & Brekke, 1981). When judging whether a man is a lawyer or an engineer, for example, people tend to give less weight than they should to statistics on the relative number of lawyers and engineers in the relevant population (Kahneman & Tversky, 1973). This error has been labeled the base rate fallacy.

Because base rate statistics are similar to incidence rate statistics, one is tempted to assume incidence rates will be underutilized as well. There are important differences between the two types of statistics, however, which render this generalization problematic. Base rate statistics indicate the frequency of a target outcome in a relevant population, while incidence rate statistics indicate the frequency of a trait or characteristic that is merely diagnostic of the target outcome. Suppose one is judging the likelihood Joan, who works in a tall building and owns a briefcase, is a lawyer. The percentage of women in Joan's building who are lawyers is a base rate statistic. The percentage of women in the building who own briefcases is an incidence rate statistic. In a criminal trial, the percentage of defendants in some relevant comparison population who are guilty is a base rate statistic,[1] while the percentage of some relevant population who possess a characteristic linking the defendant to the crime is an incidence rate statistic. Because incidence rate statistics are likely to play a different role in people's inferences than base rate statistics, people's tendency to underutilize the latter may not generalize to the former.

Research on the "pseudodiagnosticity phenomenon" (Doherty, Mynatt, Tweney, & Schiavo, 1979; Beyth-Marom & Fischhoff, 1983) has examined people's reactions to a form of statistical data more closely analogous to incidence rate statistics. In one series of studies, Beyth-Marom and Fischhoff asked people to judge the likelihood that a man, drawn from a group consisting of university professors and business executives, is a professor (rather than an executive) based on the fact he is a member of the Bears Club. These researchers were interested in how people respond to information about the percentage of professors and business executives who are "Bears." Data on the percentage of business executives who are Bears are most analogous to incidence rate statistics because they speak to the probability the man would be a Bear if he is not a professor, just as incidence rate statistics speak to the probability a defendant would possess a "matching" characteristic if he is not guilty.

When subjects in these studies were asked what information they would require to evaluate the probability that the man was a professor based on the fact he is a "Bear," most were interested primarily in knowing the percentage of professors who are Bears; only half expressed an interest in knowing the percentage of executives who are "Bears," although the two types of information are equally important (Beyth-Marom & Fischhoff, 1983, Experiment 1). Moreover, those

---

[1] Base rate statistics, when used in a trial, are sometimes called "naked statistical evidence" (Kaye, 1982). Where a person is struck by a bus of unknown ownership, evidence that a particular company operates 90% of the buses on that route is "naked statistical evidence" on who owns the bus. Where a man possessing heroin is charged with concealing an illegally imported narcotic, evidence that 98% of heroin in the U.S. is illegally imported is "naked statistical evidence" on whether the heroin possessed by the defendant was illegally imported. Courts have generally treated "naked statistical evidence" differently from incidence rate statistics. Although the majority of jurisdictions admit incidence rate statistics, courts almost universally reject "naked statistical evidence" (see, e.g., *Smith v. Rapid Transit*, 1945), though there are a few exceptions where its admissibility has been upheld (e.g., *Turner V. U.S*, 1970; *Sindell v. Abbott Labs*, 1980). For general discussions of "naked statistical evidence" see Kaye (1982), Cohen (1977), and Tribe (1971).

subjects who expressed an interest in the latter percentage often did so based on mistaken or illogical reasoning (Beyth-Marom & Fischhoff, 1983, Experiment 4). Nevertheless, when subjects were informed of the respective percentages, most subjects considered both and adjusted their beliefs in the proper direction (Beyth-Marom & Fischhoff, 1983, Experiment 5). Beyth-Marom and Fischhoff conclude that "people are much better at using [statistical] information . . . than they are at seeking it out . . . or articulating reasons for its usage" (p. 193).

Although the findings of Beyth-Marom and Fischhoff are hopeful, anecdotal evidence suggests people sometimes make serious errors when evaluating incidence rate statistics. One of the authors recently discussed the use of incidence rate statistics with a deputy district attorney. This experienced prosecutor insisted that one can determine the probability of a defendant's guilt by subtracting the incidence rate of a "matching" characteristic from one. In a case where the defendant and perpetrator match on a blood type found in 10% of the population, for example, he reasoned that there is a 10% chance the defendant would have this blood type if he were innocent and therefore concluded there is a 90% chance he is guilty. This assessment is misguided because it purports to determine the defendant's probability of guilt based solely on the associative evidence, ignoring the strength of other evidence in the case. If a prosecutor falls victim to this error, however, it is possible that jurors do as well.

The fallacy in the prosecutor's logic can best be seen if we apply his analysis to a different problem. Suppose you are asked to judge the probability a man is a lawyer based on the fact he owns a briefcase. Let us assume all lawyers own a briefcase but only one person in ten in the general population owns a briefcase. Following the prosecutor's logic, you would jump to the conclusions that there is a 90% chance the man is a lawyer. But this conclusion is obviously wrong. We know that the number of nonlawyers is many times greater than the number of lawyers. Hence, lawyers are probably outnumbered by briefcase owners who are not lawyers (and a given briefcase owner is more likely to be a nonlawyer than a lawyer). To draw conclusions about the probability the man is a lawyer based on the fact he owns a briefcase, we must consider not just the incidence rate of briefcase ownership, but also the a priori likelihood of being a lawyer. Similarly, to draw conclusions about the probability a criminal suspect is guilty based on evidence of a "match," we must consider not just the percentage of people who would match but also the a priori likelihood that the defendant in question is guilty.[2]

---

[2] Bayes' theorem may be used to calculate the amount one should revise one's prior estimate of the probability of a suspect's guilt after receiving associative evidence accompanied by incidence rate statistics (for general discussions of the use of Bayes' theorem to model legal judgments, see Kaplan, 1968; Lempert, 1977; Lempert & Saltsburg, 1977; Lindley, 1977; Schum, 1977b; Schum & Martin, 1982; Kaye, 1979). Where H and $\overline{H}$ designate the suspect's guilt and innocence respectively, and D designates associative evidence showing a match between the suspect and perpetrator on some characteristic, Bayes' theorem states:

$$p(H/D) = p(H)p(D/H)/[p(H)p(D/H) + p(\overline{H})p(D/\overline{H})].$$

The term $p(H)$ is called the prior probability and reflects one's initial estimate of the probability the suspect is guilty in light of everything that is known *before* receiving D. The term $p(H/D)$ is called the

The prosecutor's misguided judgmental strategy (which we shall call the Prosecutor's Fallacy) could lead to serious error, particularly where the other evidence in the case is weak and therefore the prior probability of guilt is low. Suppose, for example, that one initially estimates the suspect's probability of guilt to be only .20, but then receives additional evidence showing that the defendant and perpetrator match on a blood type found in 10% of the population. According to Bayes theorem, this new evidence should increase one's subjective probability of guilt to .71, not .90.[3]

There is also anecdotal evidence for a second error, which we first heard voiced by a criminal defense attorney and therefore call the Defense Attorney's Fallacy. Victims of this fallacy assume associative evidence is irrelevant, regardless of the rarity of the "matching" characteristic. They reason that associative evidence is irrelevant because it shows, at best, that the defendant and perpetrator are both members of the same large group. Suppose, for example, that the defendant and perpetrator share a blood type possessed by only 1% of the population. Victims of the fallacy reason that in a city of 1 million there would be approximately 10,000 people with this blood type. They conclude there is little if any relevance in the fact that the defendant and perpetrator both belong to such a large group. What this reasoning fails to take into account, of course, is that the great majority of people with the relevant blood type are not suspects in the case at hand. The associative evidence drastically narrows the group of people who are or could have been suspects, while failing to exclude the defendant, and is therefore highly probative, as a Bayesian analysis shows. The Defense Attorney's Fallacy is not limited to defense attorneys. Several appellate justices also appear to be victims of this fallacy (See, e.g., *People v. Robinson*, 1970). If defense attorneys and appellate justices fall victim to this fallacy, it is quite possible that some jurors do as well, thereby giving less weight to associative evidence than it warrants.

Whether people fall victim to the Prosecutor's Fallacy, the Defense Attorney's Fallacy, or neither may depend on the manner in which incidence rate statistics are presented and explained. In criminal trials, forensic experts often present information about incidence rates in terms of the conditional probability the defendant would have a particular characteristic *if he were innocent* (Jona-

posterior probability and indicates what one's revised estimate of probable guilt should be in light of everything that is known *after* receiving D. The formula indicates that the associative evidence, D, should cause one to revise one's opinion of the suspect's guilt to the extent $p(D/H)$ differs from $p(D/\overline{H})$. If one believes the suspect and perpetrator are certain to match if the suspect is guilty, $p(D/H) = 1.00$. If one believes an innocent suspect is no more likely than anyone else to possess the "matching" characteristic, $p(D/\overline{H})$ is equal to the incidence rate of the matching characteristic. This model assumes, of course, that the sole issue determining guilt and innocence is the suspect's identity as the perpetrator (rather than, say, his mental state).

[3] The prior probability of guilt, $p(H)$, is equal to .25, and because the defendant must be either guilty or innocent, $p(\overline{H}) = .75$. Because the defendent is certain to have the relevant genetic markers if he is guilty, $p(D/H) = 1.00$; and because the defendant is no more likely than anyone else to have the genetic markers if he is not guilty, $p(D/\overline{H}) = .10$, the incidence rate of the blood markers. These probabilities may be plugged into the Bayesian formula in footnote 2, allowing us to solve for $p(H/D)$, which, in this case equals .71.

kait, 1983). Where 1% of the population possess a blood type shared by the defendant and perpetrator, for example, experts often present only the conclusory statement that there is one chance in 100 that the defendant would have this blood type if we were innocent. This type of testimony seems especially likely to lead jurors to commit the Prosecutor's Fallacy. On the other hand, a defense tactic used in some actual cases is to point out that, notwithstanding its low incidence rate, the characteristic shared by the defendant and perpetrator is also also possessed by thousands of other individuals. Where 1% of the population possess a blood type shared by the defendant and perpetrator, for example, the expert might be forced to admit during cross examination that in a city of one million people, approximately 10,000 individuals would have the "rare" blood type. Statements of this type may reduce the tendency toward the Prosecutor's Fallacy but induce more errors consistent with the Defense Attorney's Fallacy.

To test these hypotheses, Experiment 1 had subjects estimate the likelihood a criminal suspect was guilty based, in part, on statistical evidence concerning the incidence rate of a characteristic shared by the defendant and perpetrator. The part of the evidence concerning the incidence rates was presented in two different ways. In one condition the forensic expert presented only the conditional probability that an innocent person would have the "matching" characteristic. In a second condition the expert stated the percentage of the population who possess the relevant characteristic and the approximate number of people who possess this characteristic in the city where the crime occurred.

EXPERIMENT 1

Method

Subjects

All subjects in this and the following experiment were volunteers from a university human subjects pool who were given extra credit in coursework as an incentive to participate. Subjects (N = 144) were run in groups of about ten in sessions lasting one half hour. Each subject was randomly assigned to one of the two experimental conditions.

Procedure

On arriving for the experiment, subjects were given a five-page packet of stimulus materials. The first page, containing instructions, stated (a) that the experiment was designed to test people's ability to draw reasonable conclusions from evidence involving probabilities, (b) that subjects would be asked to read a description of a criminal case and to indicate their estimate of the likelihood of the suspect's guilt by writing a percentage between 0 and 100, and (c) that when making these estimates subjects should disregard the concept of reasonable doubt and indicate the likelihood the suspect "really did it," rather than the likelihood a jury would convict the suspect. The experimenter reviewed these instructions with subjects and answered any questions, then subjects read the description of the criminal case.

The case involved the robbery of a liquor store by a man wearing a ski mask. The store clerk was able to describe the robber's height, weight and clothing, but could not see his face or hair. The police apprehended a suspect near the liquor store who matched the clerk's description but the suspect did not have the ski mask or the stolen money. In a trash can near where the suspect was apprehended, however, the police found the mask and the money.

At this point, subjects made an initial estimate of the probability of the suspect's guilt based only on the information they had received to that point.

Next subjects read a summary of testimony by a forensic expert who reported that samples of the suspect's hair were microscopically indistinguishable from a hair found inside the ski mask. The expert also described an empirical study that yielded data on the probability that two hairs drawn at random from different individuals would be indistinguishable. The expert's testimony was modeled on that of the actual prosecution experts in *U.S. v. Massey* (1979) and *State v. Carlson* (1978). The empirical study described in the stimulus materials was similar but not identical to that reported by Gaudette & Keeping (1974).

The experimental manipulation was the way in which the expert described the incidence rate of matching hair. In the Conditional Probability condition, the expert reported the incidence rate as a conditional probability, stating that the study indicated there "is only a two percent chance the defendant's hair would be indistinguishable from that of the perpetrator if he were innocent. . . ." In the Percentage and Number condition, the expert reported that the study indicated only 2% of people have hair that would be indistinguishable from that of the defendant and stated that in a city of 1,000,000 people there would be approximately 20,000 such individuals. Half the subjects were assigned to each condition.

After reading all of the evidence, subjects made a final judgment of the probability of the suspect's guilt.

Results

Fallacious Judgments

About one quarter of the subjects made final judgments of guilt consistent with their having fallen victim to one of the fallacies described above. Overall, 19 subjects (13.2%) were coded as victims of the Prosecutor's Fallacy because they estimated the probability of guilt to be exactly .98, which is the probability one would obtain by simply subtracting the incidence rate of the "matching" characteristic from one. Eighteen subjects (12.5%) were coded as victims of the Defense Attorney's Fallacy because their final judgment of guilt was the same as their initial judgment of guilt, which indicates they gave no weight to the associative evidence. The remaining subjects (74.3%) were coded as victims of neither fallacy because their final judgments of guilt were higher than their initial judgments (indicating they gave some weight to the associative evidence) but were less than .98.

Table 1. Number of Subjects Rendering Judgments Consistent with Fallacious Reasoning (Experiment 1)

	Mode of presentation		
Fallacy committed	Conditional probability	Percentage and number	Total
Prosecutor's Fallacy	16	3	19
Defense Attorney's Fallacy	6	12	18
Neither fallacy	50	57	107
Total	72	72	144

The manner in which the statistical information was presented significantly influenced the likelihood subjects would make judgments consistent with fallacious reasoning. Table 1 shows the number of subjects in each condition who were coded as victims of the Prosecutor's Fallacy, Defense Attorney's Fallacy, or neither fallacy. As predicted, more subjects made judgments consistent with the Prosecutor's Fallacy when the incidence rate was presented in the form of a conditional probability (22.2%) than when it was presented as a percentage and number (4.2%; χ^2 $(2, N = 144) = 10.21$, $p < .01$). Fewer subjects committed the Defense Attorney's Fallacy when the incidence rate was presented as a conditional probability (8.3%) than when it was presented as a percentage and number (16.7%), but this difference was only marginally significant ($\chi^2(2, N = 144) = 5.81$, $p < .10$).[4]

Initial and Final Judgments

After receiving only preliminary information about the arrest of a suspect, the mean judgment of the suspect's probability of guilt was .25. Subjects in the two experimental conditions had received identical preliminary information and their initial judgments did not significantly differ [Conditional Probability condition $M = .27$, Percentage and Number condition $M = .22$; $t(142) = 1.53$, $p = .13$]. Subjects' *final* judgments of probable guilt, which took into account the "match" between the suspect's and perpetrator's hair and the reported low inci-

[4] A log-likelihood ratio chi-square on the 2 × 3 table revealed a significant overall difference between the conditional probability condition and the percentage and number condition [$\chi^2(2, N = 144) = 12.26$, $p < .01$]. Multiple comparisons among the three response categories, using a simultaneous test procedure recommended by Gabriel (1966), indicated that the two conditions differed mainly in the number of subjects committing the Prosecutor's Fallacy versus neither fallacy [$\chi^2(2, N = 126) = 4.97$, $p < .01$] and in the number committing the Prosecutor's Fallacy versus the Defense Attorney's Fallacy [$\chi^2(2, N = 37) = 5.24$, $p < .10$], rather than the number committing the Defense Attorney's Fallacy versus neither fallacy [$\chi^2(2, N = 125) = .57$, n.s.]. This pattern of results allows a statistically reliable inference that the two conditions are heterogeneous with respect to the number of subjects falling in the Prosecutor's Fallacy category (compared to at least one of the other categories) [$\chi^2(2, N = 144) = 10.21$, $p < .01$] and that the two conditions differ marginally in the number of subjects falling in the Defense Attorney's Fallacy category (compared to at least one of the other categories) [$\chi^2(2, N = 144) = 5.81$, $p < .10$].

dence rate of such hair, were significantly higher than initial judgments [$M = .63$; $t(143) = 16.10$, $p < .001$, paired comparison]. More importantly, final judgments of subjects in the Conditional Probability condition ($M = .72$) were significantly higher than those of subjects in the Percentage and Number condition ($M = .53$; $t(142) = 4.34$, $p < .001$). An analysis of covariance, using initial judgments as a covariate, confirmed that this effect was significant after controlling for any initial differences [$F(1,143) = 16.14$, $p < .001$]. This finding indicates that the manner in which incidence rate statistics were presented had an important effect on subjects' judgments of probable guilt.

Comparing Subjects' Final Judgments to Model Bayesian Judgments

Between their initial and final jugments, subjects learned that the suspect's hair "matched" that of the perpetrator and that the incidence rate of such hair was only 2%. To determine whether subjects gave this information the weight it would be accorded by Bayes' theorem, subjects' final judgments were compared to "model" judgments computed by revising each subject's initial judgment in accordance with the Bayesian formula in footnote 2. For each subject, $p(D/H)$ was assumed to be 1.00 and $p(D/\overline{H})$ was assumed to be .02. Each subject's own initial judgment of probability of guilt was used as $p(H)$. These probabilities were combined using the Bayesian formula to yield a posteriori probability guilt for each subject. The model Bayesian judgments ($M = .93$) were significantly higher than subjects' final judgments [$M = .63$; $t(143) = 9.64$, $p < .001$, paired comparison]. This finding is consistent with the general tendency of people to be more conservative than Bayes' theorem when revising judgments in light of new information (Edwards, 1968).

Discussion

The results confirm suspicions, arising from anecdotal evidence, that people can make serious errors when judging guilt based on associative evidence and incidence rate statistics. About one quarter of subjects made judgments consistent with their having fallen victim to the Prosecutor's Fallacy or Defense Attorney's Fallacy. Furthermore, the number of subjects who were apparent victims of the fallacies, and mean judgments of guilt, were significantly affected by a subtle manipulation in the way incidence rate statistics were presented.

It is important to note that subjects in the two conditions did not receive different information. In both conditions subjects learned that the suspect and perpetrator matched on a characteristic found in 2% of the population. But this information was presented in ways that focused attention on different, though rather straightforward, implications of the data. Subjects in the conditional probability condition were told there was only a 2% chance the suspect would "match" if he was innocent. Presenting the data in this manner probably led more subjects to commit the Prosecutor's Fallacy because they falsely assumed the conditional probability of a "match" given innocence is the complement of the conditional probability of guilt given a "match."[5] In any case, reflection on the low probability that the match could have occurred by chance probably promoted the impression that the suspect is likely to be guilty. Subjects in the percentage and number condition were told that 2% of the population would "match" and that in a city of one million people there would be 20,000 such individuals. Presenting the data in this manner probably made subjects less likely to begin thinking about the low likelihood that the suspect would "match" if he was innocent. Instead, this presentation encourages thoughts about the large number of other individuals who also would match. This line of thinking seems more conducive to the Defense Attorney's Fallacy—and, in fact, a larger number of subjects made judgments consistent with this fallacy in the percentage and number condition, though this difference was not significant. It also creates the general impression that "a lot of people could have done it," which probably accounts for the lower estimates of probable guilt in the percentage and number condition.

Although most subjects realized that the "match" between the suspect and perpetrator was diagnostic of guilt, they may not have fully appreciated the strength of the evidence. Most subjects revised their judgments in the right direction but not by as much as Bayes' theorem would dictate given subjects' initial estimates of probability of guilt and the low incidence rate of the "matching" hair. The most obvious explanation for this apparent conservatism is that subjects gave less weight to the match between the suspect's and perpetrator's hair than this evidence deserves. One must be cautious about drawing this conclusion, however, because there are other possible explanations that are not ruled out by this design. One possibility is that subjects' initial judgments of guilt overstated their perception of the strength of the nonstatistical evidence in the case (perhaps due to a tendency to avoid the extreme lower end of the response scale). In other words, subjects' final judgments may appear conservative because a response bias in initial judgments caused the "model" Bayesian judgments to be too high, rather than because underutilization of the associative evidence caused subjects final judgments to be too low (cf. DuCharme, 1970; Slovic & Lichtenstein, 1971).

The practical significance of Experiment 1 is difficult to judge without more information. One key limitation of the experiment is that subjects were not exposed to any arguments about how much weight the statistical evidence deserved. Because jurors in actual cases may hear such arguments from attorneys or other jurors, it is important to consider how people respond to these arguments. Do they recognize the flaws in an argument for a fallacious position, or are they persuaded by fallacious reasoning? Is the impact of a fallacious argument neutralized by hearing an argument for a contrary position? Experiment 2 addresses these questions.

---

[5] During debriefing subjects were asked to explain their final judgments. Among those who judged the suspects' probability of guilt to be .98, two rationales were common. Some, like the district attorney discussed earlier, argued that a 98% chance of guilt is an obvious or direct implication of the 2% incidence rate. Others argued that if 2% of the population have hair that would "match" the perpetrator's there is only a 2% chance that someone other than the suspect committed the crime and therefore a 98% chance the suspect is guilty.

EXPERIMENT 2

Method

The procedure of Experiment 2 was similar to that of Experiment 1. Undergraduate subjects (N = 73) read a description of a murder case in which the killer's identity was unknown but the victim was known to have wounded the killer with a knife. The police find some of the killer's blood at the scene of the crime and laboratory tests indicate it is a rare type found in only one person in 100. While questioning the victim's neighbors, a detective notices that one man is wearing a bandage. Based on his overall impression of this man, the detective estimates the probability of his guilt to be .10. Later the detective receives some new evidence indicating that this suspect has the same rare blood type as the killer.

The subjects' task was to decide whether the detective should revise his estimate of the probability of the suspect's guilt in light of this new evidence and, if so, by how much. To help them make this judgment they read two arguments regarding the relevance of the blood type evidence. The "Prosecution argument" advocated the Prosecutor's Fallacy as follows:

> The blood test evidence is highly relevant. The suspect has the same blood type as the attacker. This blood type is found in only 1% of the population, so there is only a 1% chance that the blood found at the scene came from someone other than the suspect. Since there is only a 1% chance that someone *else* committed the crime, there is a 99% chance the suspect is guilty.

The "Defense argument" advocated the Defense Attorney's Fallacy:

> The evidence about blood types has very little relevance for this case. Only 1% of the population has the "rare" blood type, but in a city . . . [like the one where the crime occurred] with a population of 200,000 this blood type would be found in approximately 2000 people. Therefore the evidence merely shows that the suspect is one of 2000 people in the city who might have committed the crime. A one-in-2000 chance of guilt (based on the blood test evidence) has little relevance for proving *this* suspect is guilty.

Half of the subjects first received the Prosecution argument and then received the Defense argument (Pros–Def Condition). The remaining subjects received the arguments in reverse order (Def–Pros Condition). After reading each argument, subjects answered three questions. First, they indicated whether they believed the reasoning and logic of the argument was correct or incorrect. Then they indicated whether they thought the detective should revise his estimate of the suspect's probable guilt in light of the blood type evidence. Finally, they indicated what they thought the detective's estimate of probability of guilt should be after receiving the blood type evidence.

Notice that Experiment 2 differs from Experiment 1 in the method used to establish the suspect's initial or prior probability of guilt. In Experiment 1 subjects established the prior themselves by estimating the suspect's initial probability of guilt. They then revised their own initial estimates in light of the associative evidence. This approach left open the possibility that the conservatism of subjects' final judgments was caused by subjects' overstating their initial judgments rather than underestimating the strength of the associative evidence. In Experiment 2 the prior is established by telling subjects about the detective's initial estimate of the suspect's likelihood of guilt. Subjects are then asked how much the detective should revise this estimate in light of the "match." Asking subjects to make this judgment "in the second person" rules out a response bias in subjects' initial judgments as an explanation for conservatism.

Results and Discussion

Recognition of Fallacious Arguments

A substantial number of subjects failed to recognize that the fallacious arguments were incorrect. As Table 2 indicates, the Defense argument was more convincing than the Prosecution argument: Overall 50 subjects (68.5%) labeled the Defense argument "correct" while 21 (28.8%) labeled the Prosecution argument "correct" (McNemar $\chi^2_C(1, N = 73) = 18.23$, $p < .0001$). Only 16 subjects (22.2%) recognized that *both* arguments are incorrect. It is unclear, of course, how much subjects' perceptions of the correctness of each argument depend on the specific wordings used here.

The order in which the arguments were presented did not significantly affect ratings of correctness. The distribution of subjects across the four categories shown in Table 2 was not significantly different in the Pros–Def condition than in the Def–Pros condition ($\chi^2(3, N = 73) = 5.93$, n.s.).

Fallacious Judgments

To determine whether subjects were responding to the associative evidence in a manner consistent with fallacious reasoning, we examined responses to the questions asking whether and how much the detective should revise his estimate of the suspect's probability of guilt. These responses were divided into three categories. Judgments that the detective should increase his estimate of probable guilt to exactly .99 were coded as consistent with the Prosecutor's Fallacy because .99 is the probability one would obtain by subtracting the incidence rate of the "matching" characteristic from one. Judgments that the detective should *not* revise his estimate of probable guilt in light of the associative evidence were coded as consistent with the Defense Attorney's Fallacy. Judgments that the detective should increase his estimate of probable guilt to some level other than .99 were coded as consistent with neither fallacy. Each subject judged whether and how

Table 2. Number of Subjects Rating the Prosecution Argument and Defense Argument Correct and Incorrect (Experiment 2)

	Defense argument		
Prosecution argument	Correct	Incorrect	Total
Correct	14	7	21
Incorrect	36	16	52
Total	50	23	73

much the detective should revise his estimate at two points—once after reading each argument. Hence, the 73 subjects made a total of 146 codable responses.

Overall, responses consistent with the Prosecutor's Fallacy were rare, but responses consistent with the Defense Attorney's Fallacy were surprisingly common: only four responses (3%) were consistent with the Prosecutor's Fallacy while 82 (56%) were consistent with the Defense Attorney's Fallacy and 60 (41%) were consistent with neither fallacy. The order in which the arguments were presented did not affect the distribution of responses across the three response categories.

Seventy percent of subjects made at least one response consistent with fallacious reasoning. Three subjects (4%) made one or more judgments consistent with the Prosecutor's Fallacy, 48 subjects (66%) made one or more judgments consistent with the Defense Attorney's Fallacy, and only 22 subjects (30%) made no judgments consistent with fallacious reasoning. The order in which the arguments were presented did not significantly affect the distribution of subjects across these three categories [$\chi^2(2, N = 73) = 3.80$, n.s.].

Forty-eight percent of subjects made two responses consistent with fallacious reasoning. Thirty-four subjects (47%) made two judgments consistent with the Defense Attorney's Fallacy and one made two judgments consistent with the Prosecutor's Fallacy; none mixed fallacies.

We had expected that the likelihood subjects would respond in a manner consistent with the Prosecutor's Fallacy or the Defense Attorney's Fallacy would depend on which argument they had read. Surprisingly, the findings did not support this prediction. A between group comparison, looking at responses of subjects who had read only the first of the two arguments, revealed no significant difference between those who had read only the Prosecution argument and those who had read only the Defense argument in the distribution of the two groups across the three response categories. The number of subjects coded as victims of the Prosecutor's Fallacy, Defense Attorney's Fallacy and neither fallacy was 3, 18, and 16, respectively among those who read only the Prosecution argument, and 0, 21, and 18 among those who read only the Defense argument [$\chi^2(2, N = 73) = 3.08$, n.s.]. Nor was there any difference between these two groups in their responses after reading the second of the two arguments [$\chi^2(2, N = 73) = 1.06$, n.s.]. Regardless of which argument they read, about half or more of the subjects made judgments consistent with the Defense Attorney's Fallacy.

It is unclear why the Prosecutor's Fallacy was so much less prevalent and the Defense Attorney's Fallacy so much more prevalent in Experiment 2 than in Experiment 1. Certainly the results indicate the Defense argument was more persuasive than the Prosecution argument. Whether this finding depends on the specific arguments used here or not remains to be seen. It is interesting to note that even among subjects who had read only the Prosecution argument, nearly half responded in a manner consistent with the Defense Attorney's Fallacy. Perhaps these subjects detected something "fishy" about the Prosecution argument and therefore decided to disregard the associative evidence altogether. If this is the case, the weak Prosecution argument may actually have promoted the Defense Attorney's Fallacy.

Judgments of Probable Guilt

Although the arguments did not affect the number of subjects making judgments consistent with fallacious reasoning, they did affect subjects' judgments of probable guilt. As Table 3 shows, subjects thought the detective's subjective probability of guilt should be higher after reading the Prosecution argument (M = .31) than after reading the Defense argument (M = .24; $F(1,71) = 7.89$, $p < .01$). Order of presentation also influenced these judgments. Subjects in the Pros–Def condition thought the detective's estimate of guilt should be higher than did subjects in the Def–Pros condition, $F(1,71) = 8.44$, $p < .005$. The order effect is probably due to a simple anchoring phenomenon. Subjects who received the Prosecution argument first made higher initial judgments than subjects who received the Defense argument first. The initial judgments then served as an anchor point for subjects' second judgments. It is interesting to note, however, that a 2 × 2 analysis of variance examining the effects of type of argument (Prosecution vs. Defense) and order of presentation (Pros–Def vs. Def–Pros), revealed a significant argument by order interaction, indicating there was less variation by type of argument in the Def–Pros condition than in the Pros–Def condition [$F(1,71) = 4.43$, $p < .05$]. One way of looking at this finding is that the Defense argument, when received first, "anchored" subjects' judgments more firmly than did the Prosecution argument. This interpretation is, of course, consistent with the previously noted finding that the Defense argument was more persuasive than the Prosecution argument.

A comparison of judgments of probable guilt to a Bayesian model showed the same conservative bias that was evident in Experiment 1. According to a Bayesian analysis, the detective's subjective probability of guilt should increase from .10 to .92 after receiving the associative evidence. As Table 3 indicates, however, subjects thought the detective's revised estimate of probable guilt should be considerably lower (overall M = .28). Mean judgments were low, in part, because a substantial percentage of subjects, apparent victims of the Defense Attorney's Fallacy, indicated that the detective's estimate of probable guilt should remain at .10. Even among subjects who thought the detective's estimate should increase, however, mean judgments were well below what Bayes' theorem dictates (M = .43). These findings provide additional confirmation of subjects' conservatism

Table 3. Mean Final Estimate of Probability of Guilt (Experiment 2)

Argument		Order of presentation: Prosecution defense[a]	Order of presentation: Defense prosecution[b]
Prosecution argument	M	.42	.20
	(SD)	(.34)	(.18)
Defense argument	M	.29	.18
	(SD)	(.29)	(.14)

[a] Prosecution defense condition, n = 37.
[b] Defense–prosecution condition, n = 36.

when revising an initial judgment of probable guilt in light of associative evidence. Because subjects were indicating how much the detective should revise *his* initial estimate of probable guilt, rather than revising their own initial estimates, subjects' conservatism cannot be attributed to a tendency to avoid the lower end of the response scale when making initial judgments. Because subject's final judgments were, on average, below the midpoint of the response scale, subjects' conservatism also cannot be attributed to an artifactual tendency to avoid the upper end of the response scale (DuCharme, 1970). The most likely explanation is that subjects simply gave less weight to the associative evidence than it deserves.

GENERAL DISCUSSION

Theoretical Implications

These experiments indicate that people are not very good at drawing correct inferences from associative evidence and incidence rate statistics. They are strongly influenced by subtle and logically inconsequential differences in how the statistics are presented (Experiment 1). They are unable to see the error in crude arguments for fallacious interpretations of the evidence, and their judgments of probable guilt are strongly influenced by such arguments (Experiment 2). It appears, then, that people generally lack a clear sense of how to draw appropriate conclusions from such evidence and that, as a result, judgments based on such evidence are highly malleable.

People's responses to the evidence were far from uniform. A relatively small percentage (13% in Experiment 1; 4% in Experiment 2) gave responses consistent with the simple but erroneous judgmental strategy we have labeled the Prosecutor's Fallacy. A larger percentage (12% in Experiment 1; 66% in Experiment 2) gave responses consistent with another judgmental error, which we call the Defense Attorney's Fallacy. The remaining subjects responded in a manner that was consistent with neither fallacy but that suggested a tendency to underestimate the value of associative evidence. We will discuss each type of response in turn.

Prosecutor's Fallacy

The Prosecutor's Fallacy probably results from confusion about the implications of conditional probabilities. In the cases used in these experiments, the incidence rate statistics established the conditional probability that the suspect would "match" *if he was innocent*. Victims of the fallacy may simply have assumed that this probability is the complement of the probability the suspect would be guilty *if he matched*.

The Prosecutor's Fallacy is similar to a fallacy documented by Eddy (1982) in physicians' judgments of the probability that a hypothetical patient had a tumor. When physicians were told there is a 90% chance a particular test will be positive *if the patient has a tumor*, most of them jumped to the conclusion that there is a

90% chance the patient has a tumor *if the test result is positive*. This judgment is, of course, erroneous, except where the prior probability of a tumor and the prior probability of a positive test result are equal. But physicians made this judgment even when this condition clearly was not met. The difference between the physicians' error and the Prosecutor's Fallacy is most easily explained in formal terms: where H and $\bar{H}$ indicate that the matter being judged is true and false, respectively, and D is evidence relevant to that judgment, the physicians responded as if $p(H/D) = p(D/H)$; victims of the Prosecutor's Fallacy respond as if $p(H/D) = 1 - p(D/\bar{H})$.

The Prosecutor's Fallacy is clearly inappropriate as a general strategy for assigning weight to associative evidence because it fails to take into account the strength of other evidence in the case. Particularly where the other evidence against the defendant is weak, it can lead to errors. On the other hand, there are some circumstances in which judgments based on the Prosecutor's Fallacy will closely approximate Bayesian norms. Where the incidence rate of a "matching" characteristic is extremely low (e.g., below 1%), for example, the posterior probabilities of guilt dictated by Bayes theorem and by the Prosecutor's Fallacy will converge at the upper end of the scale and may, for practical purposes, be indistinguishable. There is also a convergence where the prior probability of guilt is near .50.[6] Whether people actually fall victim to the Prosecutor's Fallacy under all of these circumstances is, of course, speculative at this point. For present purposes it is sufficient to note that the practical consequences of the Prosecutor's Fallacy, if it occurs, are likely to be most significant where the prior probability is not close to .50 and the incidence rate is greater than .01.

Defense Attorney's Fallacy

Perhaps the most surprising finding of this research was how easily people can be persuaded to give *no* weight to associative evidence. The associative evidence presented in the two experiments was quite powerful: a match between the suspect and perpetrator on a characteristic found in only 2% (Experiment 1) or 1% (Experiment 2) of the population. According to Bayes' theorem, a person who initially thought the suspect's probability of guilt was .10 should revise that estimate upward to .85 and .93, respectively, in light of this associative evidence. Of course, one need not rely on Bayes' theorem to conclude that the Defense Attorney's Fallacy is inappropriate. It is difficult to imagine any normative model of judgment that would give no weight to associative evidence. Yet many people in

[6] In a criminal case in which the defendent and perpetrator match on a blood type found in one person in ten, for example, a victim of the Prosecutor's Fallacy would conclude that the probability of the defendant's guilt is .90. Bayes' theorem dictates a nearly identical probability (.909) when the prior probability is .50. To the extent the prior probability is lower or higher than about .50, however, the Prosecutor's Fallacy produces results that diverge from Bayesian norms. When the prior probability is .20 and the incidence rate is .10, Bayes' theorem dictates a posterior probability of .71 while the Prosecutor's Fallacy produces a posterior of .90. By contrast, when the prior is above .50, the Prosecutor's Fallacy may actually favor the defense. When the prior probability is .80 and the incidence rate is .10, the Prosecutor's Fallacy, as before, produces a posterior of .90, which is lower than the Bayesian posterior of .98.

these experiments were persuaded that because a large number of individuals other than the suspect would also "match" on the relevant characteristic, the "match" is uninformative with regard to the suspect's likelihood of guilt. As noted earlier, what this reasoning ignores is that the overwhelming majority of the other people who possess the relevant characteristic are not suspects in the case at hand. The associative evidence drastically narrows the class of people who could have comitted the crime, but fails to eliminate the very individual on whom suspicion has already focused.

The argument favoring the Defense Attorney's Fallacy was particularly persuasive. In Experiment 2, over 60% of people who heard the argument were persuaded that the associative evidence deserved no weight. Finding some way to combat this powerful fallacy is clearly an imperative for future research.

Underutilization of Associative Evidence

Among subjects who thought the associative evidence deserved some weight, final judgments of guilt still tended to be significantly lower than a Bayesian analysis suggests they should have been. Of course efforts to compare human judgments to Bayesian models are problematic (DuCharme, 1970; Slovic & Lichtenstein, 1971). But, as discussed earlier, the design and results of the two experiments appear to rule out the most obvious possible artifacts. Hence, these findings lend support to the claim that people underutilize associative evidence (Saks & Kidd, 1980).

Legal Implications

Because the present research has some important limitations, conclusions about its legal implications are best viewed as preliminary. It is unclear, for example, how much the *individual* judgmental errors documented by these studies affect the *group* decisions of juries. We do not know what happens when victims of opposing fallacies encounter each other in deliberation. Perhaps in the crucible of deliberation fallacious reasoning is detected and the jurors, as a group, adopt a more appropriate evaluation of the evidence. On the other hand, deliberation may simply cause the most persuasive line of fallacious reasoning to dominate, reinforcing the biases of the majority of jurors. These intriguing issues await further research. Another limitation of the present research is that its findings are largely based on people's estimates of probabilities rather than their decision to convict or acquit. It is important that future research confirm that the tendency to misuse associative evidence, suggested by these findings, goes beyond the articulation of numbers and actually influences the sorts of decisions juries are called upon to make.

Nevertheless, people's tendency to draw erroneous conclusions from descriptions of evidence closely modeled on evidence from actual cases is troubling. College undergraduates are unlikely to be worse at evaluating such evidence than actual jurors, so the findings suggest such evidence may well be misinterpreted and misused by juries.

As noted earlier, the legal debate over the admissibility of incidence rate statistics in connection with associative evidence stems largely from differing assumptions about the way jurors are likely to respond to such evidence. The findings of the present research cast some initial light on this issue. The finding that some people make judgments consistent with the Prosecutor's Fallacy lends support to the argument of some appellate courts and commentators (e.g., Tribe, 1971) that statistical evidence may have an exaggerated impact on the jury. On the other hand, the powerful tendency to commit the Defense Attorney's Fallacy, particularly after reading arguments, and the general tendency to make conservative judgments in light of associative evidence, suggest that underutilization of such evidence is the more serious problem.

From a legal point of view, the primary danger of admitting statistical evidence is that it will be overutilized. There is little harm in admitting it if jurors give it no more weight than it deserves. If jurors underutilize such evidence or even ignore it they may reach the wrong verdict, but they are not more likely to err if the evidence is admitted than if it is not. The major danger of admitting this evidence is therefore the possibility that juries will commit the Prosecutor's Fallacy.

Judgments consistent with the Prosecutor's Fallacy were more common when the statistical evidence was presented as a conditional probability than when it was presented as a "percentage and number" (Experiment 1). This finding suggests that courts that admit incidence rate statistics can reduce the likelihood that jurors will commit the Prosecutor's Fallacy by forbidding experts to present incidence rates as conditional probabilities, requiring them to state the incidence rate as a percentage and requiring them to provide an estimate of the number of people in the area who would also have the relevant characteristic. Lawyers' arguments may also counteract the tendency to commit the Prosecutor's Fallacy. In Experiment 2 the argument for the Defense Attorney's Fallacy proved considerably more persuasive than the argument for the Prosecutor's Fallacy. An attorney worried about jurors committing the Prosecutor's Fallacy might do well to fight fallacy with fallacy. Of course, these tactics for preventing the Prosecutor's Fallacy have a price—they make it much more likely that jurors will underutilize the associative evidence or ignore it altogether. When deciding how incidence rates should be presented to the jury, then, the key issue is not whether jurors will draw appropriate conclusions from it or not, but whether one type of error will be more likely than another.

In recent years a number of scholars have suggested that jurors be instructed in the use of "decision aids" based on Bayes' theorem in cases in which they must deal with statistical evidence (Finkelstein & Fairley, 1970; Feinberg & Kadane, 1983; Lindley, 1977). These proposals have been severely criticised by legal scholars, who argue that the proposed cure is worse than the alleged disease (e.g., Tribe, 1971; Brilmayer & Kornhauser, 1978; Callen, 1982; Cullison, 1979). Indeed, the legal community's reaction to these proposals has been so hostile that it appears unlikely any of these proposals will be adopted in the foreseeable future. Nevertheless, the present findings lend support to the claim that jurors are likely to misuse statistical evidence if they are not provided with decision aids

(Saks & Kidd, 1980). Whether jurors would evaluate statistical evidence better with decision aids than without is not an issue addressed by this research, but is a worthy topic for future study.

Another possible remedy for misuse of statistical evidence is to rely on cross-examination and arguments by the attorneys. The arguments used in Experiment 2, for example, seemed to counteract the Prosecutor's Fallacy (though they may have promoted the Defense Attorney's Fallacy). Perhaps clever lines of cross-examination or argument exist or could be developed that are effective in counteracting misuse of statistical evidence (see, e.g., Imwinkelried, 1982b). Future research might examine the way experienced attorneys actually deal with statistical evidence and might, through simulation experiments, test the effectiveness of those and other techniques.

The use of mathematical evidence is likely to increase dramatically in the near future (Jonakait, 1983) and legal professionals will increasingly face difficult choices about how to deal with it. Because their choices will turn, in part, on assumptions about the way people respond to mathematical evidence, now is an opportune time for social scientists to begin exploring this issue. Our hope is that social scientists, building on studies like those reported here, will be able to answer the key underlying behavioral questions so that lawyers and judges may base decisions about mathematical evidence on empirical data rather than unguided intuitions.

REFERENCES

Annotation (1980). Admissibility, weight and sufficiency of blood grouping tests in criminal cases. *American Law Reports, Fourth, 2*, 500.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44*, 211–233.

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

Borgida, E., & Brekke, N. (1981). The base-rate fallacy in attribution and prediction. *New directions in attribution research*, Vol. 3. Hillsdale, New Jersey: Erlbaum.

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.

Briggs, T. J. (1978). The probative value of bloodstains on clothing. *Medicine, Science and the Law, 18*, 79–83.

Brilmayer, L., & Kornhauser, L. (1978). Review: Quantitative methods and legal decision. *University of Chicago Law Review, 46*, 116–153.

Callan, C. R. (1982). Notes on a grand illusion: Some limits on the use of Bayesian theory in evidence law. *Indiana Law Journal, 57*, 1–44.

Chaperlin, K., & Howarth, P. S. (1983). Soil comparison by the density gradient method—A review and evaluation. *Forensic Science International, 23*, 161–177.

Cohen, L. J. (1977). *The probable and the provable*. London: Oxford University Press.

Cullison, A. D. (1979). Identification by probabilities and trial by mathematics (a lesson for beginners in how to be wrong with greater precision). *Houston Law Review, 6*, 471–518.

Davis, R. J., & DeHaan, J. D. (1977). A survey of men's footwear. *Journal of Forensic Science Society, 17*, 271–283.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 43*, 111–121.

Ducharme, W. M. (1970). A response bias explanation of conservatism in human inference. *Journal of Experimental Psychology, 85*, 66–74.
Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley.
Feinberg, S. E., & Kadane, J. B. (1983). The presentation of Bayesian statistical analysis in legal proceedings. *The Statistician, 32*, 88–98.
Finklestein, M. O., & Fairley, W. B. (1970). A Bayesian approach to identification evidence. *Harvard Law Review, 83*, 489–517.
Fong, W. (1973). Value of glass as evidence. *Journal of Forensic Science, 18*, 398–404.
Gabriel, K. R. (1966). Simultaneous test procedures for multiple comparisons on categorical data. *Journal of the Americal Statistical Association, 14*, 1081–1096.
Gaudette, B. D., & Keeping, E. S. (1974). An attempt at determining probabilities in human scalp hair comparison. *Journal of Forensic Science, 19*, 599–605.
Gettinby, G. (1984). An empirical approach to estimating the probability of innocently acquiring bloodstains of different ABO groups on clothing. *Journal of Forensic Science Society, 24*, 221–227.
Giannelli, P. C. (1983). *Frye v. U.S.* Background paper prepared for the national conference of lawyers and scientists. Symposium of Science and the Rules of Evidence. National Conference of Lawyers & Scientists, April 29–30.
Grunbaum, B. W., Selvin, S., Pace, N., & Black, D. M. (1978). Frequency distribution and discrimination probability of twelve protein genetic variants in human blood as functions of race, sex, and age. *Journal of Forensic Sciences, 23*, 577–587.
Imwinkelried, E. J. (1981). A new era in the evolution of scientific evidence: A primer on evaluating the weight of scientific evidence. *William & Mary Law Review, 23*, 261–287.
Imwinkelried, E. J. (1982a). Forensic hair analysis: The case against the underemployment of scientific evidence. *Washington and Lee Law Review, 39*, 41–67.
Imwinkelried, E. J. (1982b). *The methods of attacking scientific evidence*. Charlottesville, Virginia: The Michie Company.
Imwinkelried, E. J. (1983). The standard for admitting scientific evidence: A critique from the perspective of juror psychology. *Villanova Law Review, 28*, 554–571.
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, 1983*, 369–421.
Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.
Kaplan, J. (1968). Decision theory and the factfinding process. *Stanford Law Review, 20*, 1065.
Kaye, D. (1979). The law of probability and the law of the land. *University of Chicago Law Review, 47*, 34–56.
Kaye, D. (1982). The limits of the preponderance of the evidence standard: Justifiably naked statistical evidence and multiple causation. *American Bar Foundation Research Journal, 2*, 487–516.
Lempert, R. L. (1977). Modeling relevance. *Michigan Law Review, 75*, 1021–1075.
Lempert, R. & Saltzburg, S. (1977). *A modern approach to evidence*. St. Paul, Minnesota: West Publishing.
Lindley, D. V. (1977). Probability and the law. *The Statistician, 26*, 203–212.
Note. (1983). Admissibility of mathematical evidence in criminal trials. *American Criminal Law Review, 21*, 55–79.
Owens, G. W., & Smalldon, K. W. (1975). Blood and semen stains on outer clothing and shoes not related to crime: Report of a survey using presumptive tests. *Journal of Forensic Science, 20* 391–403.
Pearson, E. F., May, R. W., & Dobbs, M. D. G. (1971). Glass and paint fragments found in men's outer clothing—Report of a survey. *Journal of Forensic Science, 16*(3), 283–299.
Peterson, J. L., Mihajlovic, S., & Gilliland, M. (1984). *Forensic evidence and the police: The effects of scientific evidence on criminal investigations*. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Saferstein, R. (1977). *Criminalistics: An introduction to forensic science.* Englewood Cliffs, New Jersey: Prentice-Hall.

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

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

Schroeder, O. (1977). *Assessment of the forensic sciences profession: A legal study concerning the forensic sciences.* Washington, D.C.: Forensic Science Foundation.

Schum, D. A. (1977a). Contrast effects in inference: On the conditioning of current evidence by prior evidence. *Organizational Behavior and Human Performance, 18,* 217–253.

Schum, D. A. (1977b). The behavioral richness of cascaded inference models: Examples in jurisprudence. In Castellan, Pisoni, & Potts (Eds.), *Cognitive theory,* Vol. II. Hillsdale, New Jersey: Lawrence Erlbaum.

Schum, D. A., & DuCharme, W. M. (1971). Comments of the relationship between the impact and the reliability of evidence. *Organizational Behavior and Human Performance, 6,* 111–131.

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law and Society Review, 17,* 105–151.

Slovic, P., Fischoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39.

Slovic, P., & Lichtenstein, S. (1971). Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744.

Stoney, D. (1984). Evaluation of associative evidence: Choosing the relevant question. *Journal of the Forensic Science Society, 24,* 473–482.

Tippet, C. F., Emerson, V. J., Fereday, M. J., Lawton, F., Richardson, A., Jones, L. T., & Lampert, S. M. (1968). The evidential value of the comparison of paint flakes from sources other than vehicles. *Forensic Science Society Journal, 8,* 61–65.

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

Tversky, A., & Kahneman, D. (1982). Evidential impact of base-rates. In D. Kahneman, A. Slovic, & A Tversky (Eds.), *Judgment under uncertainty: Heuristic biases.* Cambridge: Cambridge University Press.

CASES

State v. Carlson, 267 N.W. 170 (Minn. 1978).

People v. Macedonio, 42 N.Y.2d 944, 366 N.E.2d 1355 (1977).

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

People v. McMillen, 126 Mich.App. 203, 337 N.W.2d 48 (1984).

People v. Robinson, 27 N.Y.2d 864, 265 N.E.2d 543, 317 N.Y.S.2d 19 (1970).

Sindell v. Abbott Labs, 26 Cal.3d 588, 607 P.2d 924, 163 Cal.Rptr. 132 (1980).

Smith v. Rapid Transit, 317 Mass. 469, 58 N.E.2d 754 (1945).

Turner v. U.S., 396 U.S. 398, *reh'g denied,* 397 U.S. 958 (1970).