Law and Human Behavior, Vol. 23, No. 2, 1999

Juror Reactions to DNA Evidence: Errors and Expectancies

Jason Schklar[1,2] and Shari Seidman Diamond[1]

In this paper we examine evidence for two potential descriptions of juror reactions to probabilistic DNA evidence. The error-based description posits that jurors commit systematic logical or mathematical errors when they are called upon to evaluate quantitative evidence. The expectancy-based description posits that jurors use their background knowledge and beliefs in evaluating results from scientific tests. Consistent with the error-based description, participants in our study incorrectly aggregated separately presented probabilities and afforded probabilistic evidence less weight than would be expected by applying Bayesian norms. Consistent with the expectancy-based description, participants' background beliefs about the possibility of laboratory errors and intentional tampering affected the weight participants afforded a DNA match report. We discuss potential implications of these findings for the legal system and suggest directions for future research.

DNA profiling and other kinds of scientific evidence are playing an increasing role in criminal investigations and court proceedings (National Research Council Report, 1996; Peterson, Ryan, Houlden, & Mihajlovic, 1987). As the admissibility of certain DNA profiling techniques has become less controversial (National Research Council Report, 1996), an important question remains: Are jurors able to understand the complex scientific and inherently probabilistic testimony that accompanies a DNA match report? Legal scholars (e.g., Bernstein, 1996) and other social commentators (e.g., Huber, 1991) have questioned jurors' ability to comprehend and appropriately weight scientific evidence in their judgments, and some have expressed concern that jurors may attribute a 'special aura of credibility' to scientific evidence (e.g., Imwinkelried, 1982–1983). Courts have voiced the concern that jurors may attribute an air of "mystic infallibility" to scientific evidence (e.g., *United States v. Addison*, 1974).

[1]Department of Psychology, University of Illinois at Chicago, and American Bar Foundation.
[2]Correspondence should be directed Jason Schklar, Department of Psychology (MC 285), University of Illinois at Chicago, 1007 W. Harrison, Chicago, IL 60607-7137 (email: jschklar@uic.edu).

0147-7307/99/0400-0159$16.00/1

One particular area of concern centers around the fact that DNA test results tend to be presented in a heavily quantitative manner. Some legal scholars (e.g., Tribe, 1971) have theorized that jurors are likely to be unduly influenced by overtly probabilistic evidence because it exudes an "aura of precision." Courts have echoed this concern in decisions like *People v. Collins* (1968), which raised the specter of mathematics as "a veritable sorcerer in our computerized society" who threatens to "cast a spell" over triers of fact (p. 320). More recently, courts have expressed concern that jurors might attach too much weight to the extremely small probabilities that often accompany a DNA match report (e.g., *Commonwealth v. Curnin*, 1991).

Just how jurors evaluate complex scientific evidence such as DNA profiling is an empirical question that is still only minimally informed by relevant data (Sanders, 1993). In this paper we review recent findings about how decision makers evaluate DNA and other biological profile evidence, suggest a novel way to describe these emerging findings, and report the results of a study designed to demonstrate the usefulness of this description. Finally, we discuss the implications of these findings for the legal system and suggest directions for future research.

THE DNA MATCH REPORT

In a trial involving incriminating DNA evidence, jurors are generally presented with testimony about how the crime scene DNA evidence was collected and about how the crime lab processed the evidence and tested for any matches between the various crime scene and comparison samples. When there is a report of a match between the defendant's (or sometimes the victim's) DNA profile and some crime scene DNA evidence, jurors must then consider the possibility that the match could have been declared even though the defendant was not the true source of the crime scene evidence.[3] For instance, even though no two individuals (except identical twins) have exactly the same DNA, it is possible to obtain a coincidental match report because the two main kinds of forensic DNA typing—variable-number tandem repeats (VNTR) and polymerase chain reaction (PCR)—examine only a small portion of the entire DNA sequence (for an excellent overview of the DNA profiling process, see National Research Council Report, 1992; 1996). The probability of a match being declared due to random chance is called the random match probability (RMP) and it is calculated from studies of specific allele frequencies in existing DNA databases. Depending on the rarity of the DNA profile, the size of the RMP can be vanishingly small—sometimes in the range of one in several billion (Goodman, 1992; Koehler, Chia, & Lindsay, 1995).

Human error in the DNA laboratory is another way that a match report can be made even though the defendant is not the true source of the crime scene DNA.

[3]Of course, jurors must also consider the possibility that the defendant is the true source of the crime scene evidence, but that the evidence was left in a manner consistent with his or her innocence. Consider the example (found in Koehler, 1993) of a man who is found murdered in his bed in which a DNA profile analysis of some hairs extracted from the sheets matches his wife's DNA profile. The fact that the two shared the same bed provides a completely innocuous reason for finding the wife's hairs at the crime scene.

Although some forensic scientists have argued that it is impossible to obtain a false-positive match using current DNA technology (for some examples of such in-court testimony, see Koehler, 1993; Thompson, 1993), a review of the few published proficiency test reports belie this claim (e.g., Koehler et al., 1995; National Research Council Report, 1992, 1996; Thompson, 1993). The probability that a match report was declared due to a human error in the DNA lab is known as the laboratory error rate (LE).

HOW DO JURORS EVALUATE DNA AND OTHER BIOLOGICAL PROFILE EVIDENCE?

A small body of research has examined how decision makers (undergraduates, adult jury-eligible community members, and adults summoned for jury duty who are waiting to be selected) evaluate probability estimates associated with biological profile evidence such as blood typing, hair fibers, and DNA. In one of five published studies, participants overweighted the evidence (Koehler et al., 1995) and in the remaining four they underweighted it (Faigman & Baglioni, 1988; Goodman, 1992; Smith, Penrod, Otto, & Park, 1996; Thompson & Schumann, 1987). Although the five studies were similar in a number of respects (i.e., they all involved the presentation of incriminating probabilistic evidence at a criminal trial), some distinguishing characteristics can be identified that may account for the apparently discrepant pattern of findings.

Perhaps the most important difference among the studies was the normative baseline against which decision makers' responses were compared (for a review of some of the other differences among the five published studies, see Schklar, 1996, July). In four of the studies Bayes' theorem was used to estimate how much participants should be influenced by a new piece of probabilistic evidence (Faigman & Baglioni, 1988; Goodman, 1992; Smith et al., 1996; Thompson & Schumann, 1987). Generally, this involved three steps: First, immediately after receiving some non-DNA evidence and immediately before receiving the probabilistic evidence, participants reported their estimate of the likelihood that the defendant committed the crime (prior probability of guilt estimate). Second, participants were presented with the incriminatory probabilistic evidence (which was always an RMP-type estimate in these studies, never LE). And third, participants reported their revised subjective belief that the defendant committed the crime given the DNA match report (posterior probability of guilt estimate).

The degree to which decision makers should have been influenced by the biological match report was calculated by applying Bayes' theorem to participants' prior probability of guilt estimates and the RMP estimate. This normative estimate was compared with participants' actual posterior probability of guilt estimates to examine whether they had over-, under-, or appropriately weighted the probabilistic evidence. In all four of the Bayesian updating studies, participants gave the probabilistic evidence less weight than Bayes' theorem specified was appropriate.

Koehler et al. (1995) used a different normative yardstick to measure decision makers' use of probabilistic evidence. These researchers examined whether laypersons could combine probability estimates appropriately in cases in which one estimate was extremely small and the other estimate was many orders of magnitude larger. Study participants were presented with zero, one, or two estimates of the probability that a match was declared between some crime scene evidence and the defendant even though the defendant was not the true source of the crime scene evidence. Participants in the two separate estimates condition received both an extremely small RMP estimate (1 in a billion) and a much larger LE estimate (2 in 100).[4] Notice that given the two separate estimates, the probability that a match was declared due to either random chance or human error is equal to the sum of the probability of either one occurring (or, 1 in a billion plus 2 in 100, which equals approximately 2 in 100) minus the probability of their joint occurrence (which, assuming the two events occur independently, is minuscule—2 in 100 times 1 in a billion). In essence, the 2 in 100 LE figure "dwarfs" the RMP estimate of 1 in a billion, rendering it almost inconsequential.

If participants understood that the much larger LE estimate essentially "dwarfed" the much smaller RMP estimate, those who received both estimates should have (a) convicted the defendant as often as their counterparts who received a single normatively combined estimate of RMP and LE of 2 in 100 (and who were not told the size of the separate RMP and LE estimates), and (b) convicted the defendant less often than participants who received only the much smaller RMP estimate. In contrast, Koehler et al. (1995) found that participants who received separate RMP and LE estimates convicted the defendant as often as—and, in a second study, slightly more often than—participants who received a single RMP estimate of 1 in a billion, and significantly more often than participants who received a single normatively combined estimate of RMP and LE of 2 in 100. Koehler et al. (1995) concluded that participants who received separate RMP and LE estimates were unduly influenced by the extremely small—yet essentially irrelevant—RMP of 1 in a billion.

DESCRIBING JURORS' (MIS)USE OF PROBABILISTIC DNA EVIDENCE

The Error-Based Description

Traditionally, the results of these five studies have been described in terms of what we call "error-based" descriptions. The error-based description posits that jurors are imperfect fact finders who predictably misuse probabilistic DNA evidence because they commit logical or mathematical errors, such as incorrectly aggregating separately presented probabilities and affording probabilistic evidence less weight than Bayesian norms would indicate. In this section we present a useful conceptualization of the kinds of errors that jurors may commit when they evaluate probabilistic DNA evidence and we review the few existing efforts to correct these errors.

[4] Actually, participants were presented with an LE of either 2 in 100 or 1 in 1,000. Because size of the LE did not affect conviction rates, we discuss the results using the 2 in 100 estimate.

Misaggregation and Misperception

The fact that human performance deviates from the rational yardstick (no matter which one we use) is not surprising: Social psychologists have often exposed foibles in human reasoning (Funder, 1987). One way of explaining jurors' misuse of probabilistic DNA evidence is in terms of errors in probabilistic inference. DuCharme (1970) provides a useful conceptualization of two independently occurring errors—misaggregation and misperception—that people may commit when evaluating probabilistic evidence.

The misaggregation error arises when a person's subjective belief in the validity of a hypothesis is not updated to the extent that is logically warranted based on prior held beliefs and the probative value of a new piece of probabilistic evidence. The finding that participants in the Faigman and Baglioni (1988), Goodman (1992), Smith et al. (1996), and Thompson and Schumann (1987) studies undervalue probabilistic evidence when compared to Bayesian norms can be seen as evidence of the misaggregation error. Participants were less impressed by the probabilistic evidence than was logically warranted, a finding that is consistent with Saks and Kidd's (1980) review of the social psychology literature on the processing of quantitative information.

The misperception error arises when a person misjudges the probative value of a new piece of probabilistic evidence and consequently affords it either too much or too little weight in his or her decision. Koehler et al. (1995) offer two explanations for why participants in their study afforded extremely small probabilities too much weight in their decisions: the vividness hypothesis (Nisbett & Ross, 1980) and the averaging strategy hypothesis (originally suggested by Lempert, 1993). Consistent with the vividness hypothesis, participants may have misperceived the probative value of the DNA test results because they were unduly influenced by the extremely small (i.e., more vivid and memorable) RMP estimate of 1 in a billion than the much larger (i.e., more pallid and less memorable) LE estimate of 2 in 100; or, consistent with the averaging strategy hypothesis, participants may have averaged the denominators of the separate RMP and LE fractions together.[5] Both of these explanations can be conceptualized as misperception errors.

DuCharme (1970) also noted that misperception and misaggregation are not mutually exclusive, and that these two errors may occur in tandem. Thus, even though Koehler et al. (1995) found that participants afforded the extremely small RMP information more weight than was normatively appropriate (evidence of the misperception error), they still may have underweighted this evidence overall when compared to Bayesian predictions (evidence of the misaggregation error). Although Koehler et al. (1995) did not collect prior and posterior probability of guilt esti-

[5]Note the difference between averaging the denominator of 1 in a billion and 2 in 100 and taking an average of the two full fractions is quite substantial. People who average the denominators would arrive at a combined estimate of approximately 1 in 500 million (.000000002), which gives far too much weight to the 1 in a billion probability estimate. On the other hand, people who average the two full fractions would arrive at a combined estimate of approximately 1 in 100 (.0100000005), which is much closer to the normatively correct 2 in 100 probability that the match could have occurred due to either RMP or LE.

mates—precluding a Bayesian analysis of their data—participants likely underweighted the probabilistic evidence to at least some degree because in even the most prejudicial experimental condition (where jurors were given an RMP of 1 in a billion and no LE information) the conviction rate was only about 44% in Study 1 and 54% in Study 2. It seems unlikely that posterior probability of guilt estimates were exceptionally high in these conditions, or the conviction rate would have been much higher.

Efforts to Correct Errors

Efforts to address the misaggregation error by helping jurors perform as better Bayesians have not been successful. In a study by Faigman and Baglioni (1988), participants received an expert's explanation of how to weigh the probabilistic evidence according to Bayes' theorem. Participants generally misunderstood and ignored this testimony and ended up underutilizing the probabilistic evidence as compared to Bayesian norms. In a more recent study by Smith et al. (1996), some participants received Bayesian training by an expert witness, while the rest did not. No difference was obtained between the two groups in terms of performance relative to Bayesian norms, and in all cases participants tended to underweight the probabilistic evidence.

No published efforts have addressed how the misperception error might be attenuated. Participants in the Koehler et al. (1995) study were never told how they should combine separate RMP and LE estimates, so whether or not some instruction would have enabled them to combine the two estimates appropriately remains an unanswered question. As Wilson and Brekke (1994) point out, to engage in error correction a person must be aware of it, motivated to correct for it, and able to correct for it. Arguably, participants in the Koehler et al. (1995) study were *motivated* to correct for an error that hurt the defendant because of defendant protection norms (see MacCoun and Kerr, 1988, for a meta-analysis and review). However, it is unlikely that participants were either *aware* of the fact that they were erroneously overweighting the RMP estimate of 1 in a billion, or *able* to combine the RMP and LE estimates correctly based on their limited knowledge of DNA testing procedures and the relevant probability theory.

The Expectancy-Based Description

In addition to making logical or mathematical errors, we propose that jurors may also be influenced by their background beliefs about the possibility of laboratory errors and intentional tampering that they bring with them into the jury box. The implication of this "expectancy-based" description is that juror reactions to probabilistic DNA evidence may not necessarily represent departures from a rational baseline (i.e., errors), but may instead represent their effort to incorporate their beliefs about the "real world," both accurate and inaccurate, into their decisions.

Laboratory Errors

In all five prior studies analyzing responses to probability information about biological profile evidence, participants received information about the probability of the match occurring by random chance (i.e., the RMP). Only in the Koehler et al. (1995) study, however, were participants also provided with information about the reliability of the test results (i.e., the LE). Do jurors simply assume that DNA test results are error-free when they are provided with no LE information?

Although no published study has reported jurors' naive expectancies of how likely it was that a DNA match report could have resulted from either random chance or a laboratory error, some evidence indicates that people think human errors in the DNA lab are more likely than proficiency test results have revealed. In a study of naive perceptions of how accurate and error-prone DNA test results are, Schklar (1996, March) found that undergraduates estimated the probability that a match report occurred due to LE to be about 1 in 10. The modal naive estimate of LE was also 1 in 10 for participants in a pilot study we conducted using the original Koehler et al. (1995) stimulus materials. These estimates of LE are between 5 and 100 times larger than DNA crime lab proficiency testing has revealed (Koehler et al., 1995; National Research Council Report, 1996).

The notion that expectancies about LE might influence juror decisions casts previous empirical findings in a new light. First, it suggests a need to reconsider the apparent underweighting of the probability estimates in the Bayesian updating studies: Were participants underweighting the RMP information, or were they considering the possibility that the lab tests were not error-free? It is possible that at least part of the reason participants did not appear to adjust their posterior probability of guilty estimates to the extent that Bayes' theorem would predict is because they used their own naive estimates of LE in their decisions and those estimates were larger than the RMP estimates they received.[6]

Second, uninformed expectancies about LE may partially explain the Koehler et al. (1995, Study 2) results in which participants who received both RMP (of 1 in a billion) and LE (of 2 in 100) estimates convicted the defendant slightly more often than their counterparts who had only received an RMP estimate (and no LE). Perhaps providing participants with an LE estimate had the unintended effect of *increasing* confidence in the DNA profiling results because the likelihood that a

[6]This can be illustrated using Bayes' theorem. The following equation (found in Hays, 1994) can be used to calculate the normative posterior probability of guilt estimate $p(G|M)$ based on several parameters: $p(G)$ = prior probability the defendant is guilty, $p(\sim G)$ = prior probability the defendant is not guilty, $p(M|G)$ = probability that a match will be reported given the defendant is guilty, and $p(M|\sim G)$ = probability that a match will be reported given the defendant is not guilty:

$$p(G|M) = p(G)\frac{p(M|G)}{p(G)p(M|G) + p(\sim G)p(M|\sim G)} \tag{1}$$

Given that RMP and LE are both estimates of $p(M|\sim G)$, located in the denominator of this equation, the normatively appropriate posterior probability of guilt estimate $p(G|M)$ gets smaller as estimates of $p(M|\sim G)$ increase.

human error in the laboratory occurred was smaller than what they had naively expected!

Intentional Tampering

Jurors may also consider the possibility that the match could have been declared between the defendant and some crime scene evidence due to intentional tampering by the police, prosecutor, or criminalist. Although it is impossible to estimate the frequency with which this kind of conduct occurs, documented cases of intentional tampering do exist (see, e.g., Connors, Lundregan, Miller, & McEwen, 1996). In none of the previous studies were participants' expectancies about the likelihood of intentional tampering assessed, so it is unknown whether these background beliefs influenced the amount of weight participants afforded the probabilistic evidence.

Source of the Probability Estimate

Another issue that has not been addressed is whether jurors consider the source of the probability estimate (RMP or LE) in addition to its size (1 in a billion or 2 in 100). Consider two different sets of DNA test results that yield an overall probability of 2 in 100 that the match was declared due to either random chance or human error in the laboratory. In the first test, the RMP is 1 in a billion and the LE is 2 in 100 (as in the original Koehler et al., 1995, study). In the second test, the RMP is 2 in 100 and the LE is 1 in a billion. Even though, mathematically speaking, the probability that a match was declared when the defendant was not the true source of the crime scene DNA evidence is the same for both tests, do jurors see both sets of test results in the same light? Or, do jurors psychologically differentiate between a test that is apparently extremely discriminating yet error-prone and a test that is apparently less discriminating yet virtually error-free?

THE CURRENT STUDY: OVERVIEW AND HYPOTHESES

Overview

We designed this study to test several hypotheses about the extent to which reactions to DNA evidence are guided by both errors and expectancies. Decision makers received a brief case description that included the probabilities associated with a DNA match report, were provided with minimal instructions on the law,[7] and did not deliberate to a group verdict. As "Stage One" research (Diamond, 1997), this experiment represents a preliminary investigation using college students

[7]They were told to find guilt only if the evidence proved guilt beyond a reasonable doubt. Although judges typically do not give jurors any specific instructions on how to handle DNA evidence, the National Research Council Report (1996) has described circumstances in which such instructions would be advisable (see pp. 198–199, notes 93 and 95).

to study some basic cognitive and information processing patterns that may reflect the behavior of jurors and other decision makers.

Building on the Koehler et al. (1995) design, we assessed decision makers' verdict choices in response to various presentations of probabilistic RMP and LE estimates. However, we also made two crucial additions. First, we added an experimental condition in which decision makers were instructed how to combine separately presented RMP and LE estimates correctly. Second, we partially crossed the size of the probability estimates (1 in a billion or 2 in 100) with the type of probability estimate (RMP or LE) in order to examine whether mathematically equivalent DNA test results were viewed as psychologically equivalent.

We also used a series of measures to analyze the processes that underlie verdict preferences. Consistent with the Bayesian updating studies, we assessed decision makers' subjective probability estimates. In addition, we also asked decision makers to estimate several other parameters (such as the probability that the DNA test results could have occurred due to intentional tampering) that we could include in our Bayesian analysis in order to better model their decisions.

Because our study involves predictions about both decision makers' verdict preferences and subjective probability estimates, we generated two sets of research hypotheses to describe their use of DNA evidence. The first set of hypotheses focuses on how the presentation of DNA test results affects verdict preferences. The second set of hypotheses focuses on decision makers' subjective probability estimates. Taken together, these hypotheses shed light on how reactions to DNA evidence are shaped by both errors and expectancies.

Verdict Preference Hypotheses

Hypothesis 1: One in a Billion Estimate Versus Combined Estimate of Two in a Hundred. Decision makers who receive a probability estimate of 1 in a billion[8] (whether or not it is accompanied by a second probability estimate of 2 in 100) will convict the defendant more often than decision makers who receive a single normatively combined probability estimate of 2 in 100.

Hypothesis 2: One in a Billion Alone Versus with Two in a Hundred. Decision makers who receive two separate probability estimates of 1 in a billion and 2 in 100—with no combination instructions—will misperceive the probative value of the match report and convict the defendant as often as if they had only received a single estimate of 1 in a billion.

Hypothesis 3: Instruction Effect. Decision makers who receive two separate probability estimates of 1 in a billion and 2 in 100 and who are provided with combination instructions will correctly perceive the probative value of the DNA match report and will convict the defendant less often than decision makers who receive two separate probability estimates with no combination instructions and decision makers who receive only a single probability estimate of 1 in a billion.

[8]The first three hypotheses concern the size (1 in a billion or 2 in 100) and number (one or two) of the probabilities decision makers receive. These predictions are the same whether the RMP or the LE was the source of the 1 in a billion probability.

Hypothesis 4: Source Effect. Will decision makers view a test that appears to be extremely discriminating, yet is error-prone (RMP of 1 in a billion and LE of 2 in 100) differently than a test that appears to be not quite as discriminating, but is virtually error-free (RMP of 2 in 100 and LE of 1 in a billion), even though the probability that a match was declared due to either random chance or laboratory error is the same in both cases (about 2 in 100)? This hypothesis is presented in the form of a question because, as noted above, although the two versions are normatively equivalent, little is known about whether decision makers will perceive them to be equivalent.

Hypothesis 5: Source Omission Effect. Decision makers who receive an LE estimate of 1 in a billion (and no RMP estimate) will convict the defendant more often than decision makers who receive an RMP estimate of 1 in a billion (and no LE estimate). This prediction is based on the assumption that decision makers will fill in the "missing information" with the belief that laboratory errors are a much more likely source of incorrect match reports than random chance.

Subjective Probability Hypotheses

Hypothesis 6: Memorability of the Probability Estimates. If the vividness hypothesis explains why people pay undue attention to extremely small probability estimates, decision makers will find the extremely small probability estimate (1 in a billion) more memorable than the larger probability estimate (2 in 100).

Hypothesis 7: Combination Strategies. If the averaging strategy hypothesis explains why people incorrectly combine separately presented probability estimates, then decision makers' estimates of the probability that the match report was declared due to either RMP or LE will be equal to the average of the denominator of the two separate estimates.

Hypothesis 8: Underweighting of Probabilistic Evidence. Decision makers' posterior probability of guilt estimates will be lower than Bayesian norms given their prior probability of guilt estimates and the expert-provided probability estimates.

Hypothesis 9: Personal Estimates. The gap between decision makers' posterior probability of guilt estimates and Bayesian norms will be attenuated when decision makers' own personal estimates of RMP, LE, and intentional tampering are included in the Bayesian model.

Method

Participants

Two hundred and nineteen jury-eligible (18 years of age or older, U.S. citizen, and fluent in the English language) undergraduate psychology students participated in this study as part of a course requirement. Participants were deemed "fluent" if they had indicated that English was either the language they spoke most often or the main language spoken in their home. Fifty-three percent of the participants were women, the mean age was 19.5 years ($SD = 2.8$), and the ethnicity distribution was 51% White, 17% Asian, 16% Hispanic/Latino, 11% African American, and 4% other. Participant gender, age, and ethnicity were not systematically related to the main dependent measures of interest.

Procedure

Students participated in sessions of 7–15 people, and were randomly assigned to experimental conditions within each session. All participants were told that they would be reading a two page crime scenario vignette in which the defendant was on trial for sexual assault. Participants learned that the victim did not previously know the defendant and could not identify him in a lineup because he had been wearing a mask during the attack. The defendant was originally stopped by police for making an illegal U-turn on the same night as the attack, and subsequently arrested for carrying an illegal knife. The defendant became a suspect in the sexual assault because he was arrested in the vicinity of the attack and had a weak alibi about his previous whereabouts.

An expert witness (a criminalist) then testified that DNA extracted from semen stains found on the victim matched the defendant's own DNA. Participants then read one of seven versions of the criminalist's expert testimony (described below) which conveyed information about the RMP, the LE, or both, depending on the experimental condition. Finally, participants filled out a questionnaire that included a variety of manipulation checks and dependent measures.

Independent Variables

Criminalist's Expert Testimony. There are seven versions of the criminalist's expert testimony. Information about the source of the DNA match report (RMP or LE) was partially crossed with information about the size of the probability estimates (1 in a billion, 2 in 100, or not given), yielding an incomplete factorial design. In the single-estimate (SE) condition, half of the participants learned that the RMP was 1 in a billion and were given no LE information, and the other half learned that the LE was 1 in a billion and were given no RMP information. In the two-separate-estimates (TSE) condition, half of the participants learned that the RMP was 1 in a billion and the LE was 2 in 100, and the other half learned that the RMP was 2 in 100 and the LE was 1 in a billion. The two-separate-estimates plus combination-instructions (TSECI) condition was identical to the TSE condition except that the criminalist also explained how to combine the RMP and LE estimates mathematically. Participants learned that by summing the two estimates and subtracting the probability of their joint occurrence they should end up with a combined estimate of "about 2 in 100." Finally, in the single-combined-estimate (SCE) condition, participants were not given separate estimates of RMP and LE, but were instead told that the probability of either one occurring was 2 in 100.

Prior Probability Probe. Immediately prior to receiving the DNA test results, half of the participants were asked to quantify their estimate of the probability that the defendant had committed the crime (a prior probability of guilt estimate). We included this manipulation in order to assess whether asking participants to quantify their judgments early on in the process would systematically influence any of the main dependent measures (see Faigman and Baglioni, 1988, and Thompson and Schumann, 1987, for a more detailed discussion of this issue). This manipulation increased the size of the experimental design from 7 to 14 cells.[9]

Dependent Variables

Verdict Preferences. Participants indicated their dichotomous verdict choice (guilty or not guilty) after being told that they should "find against Steven Murphy [the defendant] only if the evidence convinces you 'beyond a reasonable doubt' that Steven Murphy is guilty of this crime."

Subjective Probabilities. Participants indicated their estimate of the probability that the defendant had committed the crime (a posterior probability of guilt estimate).

Participants also answered several questions about the probabilities they had (or had not) been presented with. In addition to testing whether participants could recall the expert-provided probabilities, we asked participants to indicate what they "really" thought these probabilities were, based on all the evidence they had heard. This provided participants with a chance to express whether or not they believed the probabilities that the expert had presented (for example, did participants really believe that it was possible for a laboratory to have an error rate as low as 1 in a billion?)

We also assessed participants' uninformed expectancies about the size of RMP, LE, or the probability that either one had occurred by asking participants who had not received this information to give their best estimate of how probable they thought these events were.

Finally, participants indicated their expectancies about the probability that the match report could have occurred due to intentional tampering with the DNA evidence by the police, prosecutor, or criminalist.

Manipulation Checks

After participants had indicated their verdict preferences and subjective probability estimates, they were given a brief multiple-choice test about the definitions of RMP and LE, the size of the probability estimates they were presented with, and how they were instructed to combine these two estimates. Options of both "not mentioned" and "cannot remember" were included for each question so that it was possible to differentiate between participants who forgot the information they were exposed to and those who correctly noted that they had not been exposed to the information.

[9]Copies of the experimental materials are available from the first author upon request.

Results

Manipulation Checks

Probability Estimates. Overall, participants in the SE, TSE, and TSECI conditions correctly recognized the RMP and LE estimates they were presented with (78% correct for both). Similarly, 71% (20 of the 28) of the participants in the SCE condition correctly recognized that the expert's estimate of the probability that either random chance or human error could have produced the match report was 2 in 100. Importantly, the pattern of verdicts did not change substantially when only participants who correctly recognized the probabilities they had received ($N = 152$) were considered (two minor exceptions are reported in the Verdict Preferences section).

Combination Instructions. The presentation of a simple combination instruction did convey to some participants information about how to combine separate RMP and LE estimates. While no participants in the TSE condition reported that they had been told to sum the two estimates, 41% of participants in the TSECI condition reported that they had been told to sum the two estimates, χ^2 $(N = 131) = 33.96, p < .05$. Moreover, 81% of participants in the TSECI condition correctly reported that they had been told that there was a 2 in 100 probability that the match report occurred due to either random chance or human error in the lab, even though less than half of them remembered how the 2 in 100 combined probability estimate was computed.

Verdict Preference Data (Hypotheses 1–5)

To test the verdict preference hypotheses, we analyzed conviction rates using logistic regression and a series of five planned orthogonal contrasts (see Table 1). We also assessed whether our verdict patterns might have been qualified by either the prior probability probe or a potential higher order interaction. As it turns out, the prior probability probe exerted neither a main nor any interactive effects on verdict preferences (all Walds < 2.97, *ns*). Accordingly, Figure 1 shows the verdict preference data collapsed across prior probability probe.[10]

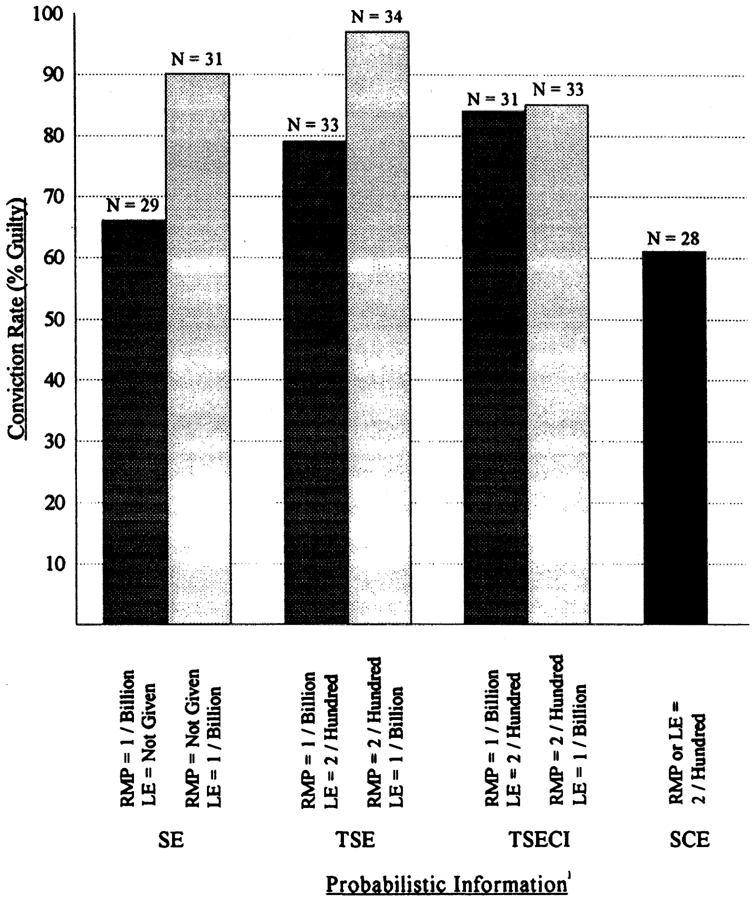
Hypothesis 1: One in a Billion Estimate Versus Combined Estimate of Two in a Hundred. Consistent with Hypothesis 1, participants convicted the defendant significantly more often when provided with an extremely small probability estimate (78%, 88%, and 84% in the SE, TSE, and TSECI conditions, respectively) than when they were given only a larger normatively combined

[10]One probe effect did emerge when we ran the logistic regression analysis using only those participants who correctly recalled the probabilities they had received. Participants who received a prior probability probe convicted the defendant more often (87%) than participants who did not receive a prior probability probe (74%).

Table 1. Planned Contrasts for Verdict Preference Measures

	SE				TSE				TSECI					
	RMP		LE		RMP		LE		RMP		LE		SCE	
Main contrasts	P	NP	P	NP	P	NP	P	NP	P	NP	P	NP	P	NP
1. Hypothesis 1 (billion v. combined)	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	–6	–6
2. Hypothesis 2 (billion alone v. billion with 2 in 100)	+1	+1	+1	+1	–1	–1	–1	–1	0	0	0	0	0	0
3. Hypothesis 3 (instruction effect)	+1	+1	+1	+1	+1	+1	+1	+1	–2	–2	–2	–2	0	0
4. Hypothesis 4 (source effect)	0	0	0	0	+1	+1	–1	–1	+1	+1	–1	–1	0	0
5. Hypothesis 5 (source omission effect)	+1	+1	–1	–1	0	0	0	0	0	0	0	0	0	0

Note: Probabilistic information: SE = single estimate (either LE or RMP); TSE = two separate estimates (both LE and RMP); TSECI = two separate estimates (both LE and RMP) plus combination instructions; SCE = single combined estimate. The estimate given as 1 in a billion was either the RMP (random match probability) or the LE (laboratory error rate). Further, a prior probe was either given (P) or not given (NP).



[1]
SE = Single Estimate
TSE = Two Separate Estimates
TSECI = Two Separate Estimates Plus Combination Instructions
SCE = Single Combined Estimate

Fig. 1. Conviction rate collapsed across prior probability probe.

probability estimate (61% in the SCE condition), Wald = 7.87, $p < .05$ (contrast 1).[11]

Hypothesis 2: One in a Billion Alone Versus with Two in a Hundred. Although Hypothesis 2 specified that there would be no difference between these two con-

[11]When we ran the logistic regression analysis using only those participants who correctly recalled the probabilities they had received, contrast 1 was reduced to marginal significance ($p = .11$). This occurred even though the conviction rates for this smaller group ($N = 152$) in the SE (75%), TSE (89%), TSECI (84%), and SCE (65%) conditions were almost identical to the conviction rates reported in the text for the full sample ($N = 219$), suggesting that the failure of this contrast to reach statistical significance was due to decreased power.

ditions, participants in the TSE condition convicted the defendant significantly *more* often (88%) than participants in the SE condition (78%), Wald = 10.65, $p < .05$ (contrast 2). Even though this finding is still evidence of the misperception error (i.e., participants incorrectly combined the two separate probability estimates), it is somewhat inconsistent with the vividness and averaging strategy hypotheses. According to the vividness hypothesis, if participants in the TSE condition were unduly influenced by extremely small (vivid) probabilities and ignored larger (pallid) probabilities, they should have convicted the defendant as often as participants in the SE condition who received only the extremely small probability estimate. According to the averaging strategy hypothesis, if participants in the TSE condition were combining the two estimates by averaging the denominators of each component, they should have convicted the defendant slightly less often than participants in the SE condition.

Hypothesis 3: Instruction Effect. A simple explanation of how to combine the RMP and LE probability estimates correctly did not ameliorate the misperception error. Participants in the TSECI condition (84%) convicted the defendant as often as participants in the SE (78%) and TSE (88%) conditions, Wald = 1.86, *ns* (contrast 3), a finding that is inconsistent with Hypothesis 3. It should be noted that the failure of the combination instruction to lower conviction rates was not simply due to the fact that participants failed to absorb the information they were given. Conviction rates remained the same even when we included only the responses of those participants who remembered the expert's combination instructions.

Hypothesis 4: Source Effect. Participants who received extremely small RMP and much larger LE estimates convicted the defendant significantly less often (81%) than participants who received extremely small LE estimates and much larger RMP estimates (91%), Wald = 14.50, $p < .05$ (contrast 4). This finding suggests that participants evaluated the results of an apparently discriminating, yet error-prone test (RMP of 1 in a billion and LE of 2 in 100) less favorably than the results of an apparently less discriminating, yet virtually error-free test (RMP of 2 in 100 and LE of 1 in a billion). Note that this finding is at odds with both the vividness and the averaging strategy explanations of the misperception error. According to the vividness hypothesis, participants should have been equally impressed by the extremely small (vivid) 1 in a billion probability estimate regardless of whether it was an estimate of RMP or LE. Likewise, if participants were simply averaging the denominators of the two probability estimates, they should have given the same weight to the apparently discriminating, yet error-prone test that they gave to the apparently less discriminating, yet virtually error-free test.

Hypothesis 5: Source Omission Effect. Consistent with Hypothesis 5, participants' expectancies about missing LE and RMP information guided their use of the DNA test results. Participants convicted the defendant significantly more often when they received extremely small LE estimates and no RMP information (90%) than when they received extremely small RMP estimates and no LE information (66%), Wald = 4.80, $p < .05$ (contrast 5). In other words, participants were less confident in the test results when their concerns about LE went unaddressed than when their concerns about RMP went unaddressed.

Summary of Verdict Preference Data. Consistent with an errors description, participants appeared to misperceive the probative value of a DNA test result when presented with separate estimates of RMP and LE. Further, the misperception error was not attenuated when participants were provided with instructions about how to combine separate RMP and LE estimates correctly. However, neither the vividness hypothesis nor the averaging strategy hypothesis appears to offer a satisfactory explanation of the misperception error.

Consistent with an expectancies description, conviction rates were highest when LE was 1 in a billion, next highest when LE was 2 in 100, and lowest when no LE information was provided (see Figure 1). One potential explanation of this finding is that participants naively expected LE to be more likely than what the proficiency test data revealed. As a result, participants were more skeptical of the DNA test results when these expectancies went unchallenged (i.e., no LE information was provided) than when the expert criminalist challenged these expectancies by providing an LE estimate that was smaller than what participants had naively assumed.

Subjective Probability Data (Hypotheses 6–9)

Hypothesis 6: Memorability of the Probability Estimates. To test whether the vividness hypothesis might explain the pattern of conviction rates, we analyzed how well participants who had received two separate probability estimates (i.e., in the TSE and TSECI conditions, $N = 131$) remembered both the extremely small (i.e., vivid) and much larger (i.e., pallid) figures. Collapsed across type of estimate (RMP or LE), participants were actually less likely to remember the extremely *small* probability (74% correct) than the *larger* probability (85% correct). Seven percent were correct only on the small probability, while 18% were correct only on the large, McNemar test approximate χ^2 (df $= 1$, $N = 32$) $= 5.28$, $p < .05$. Because this difference is precisely opposite to the one that Hypothesis 6 predicted, the pattern provides no support for claims that vividness is responsible for the misperception error.

Hypothesis 7: Combination Strategies. We categorized participants in the TSE condition ($N = 67$) according to whether their own estimate of the probability that the match was declared due to random chance or human error in the laboratory was most similar to the probability they would have obtained had they actually summed (and subtracted the joint probability), averaged (either the denominators or the full fractions), or multiplied the expert-provided RMP and LE estimates together. Consistent with Hypothesis 7, participants most frequently appeared to average the denominators of the separate RMP and LE estimates together (44% did so). Other strategies included summing them and subtracting the joint probability of their occurrence (33%), multiplying them together (14%), and averaging the two terms (8%). Note that although one third of the participants appeared to combine the two separate estimates correctly (by summing them and subtracting the joint probability of their occurrence), the remaining two-thirds of the participants engaged in some form of erroneous combination strategy (averaging or multiplication).

Hypothesis 8: Underweighting of Probabilistic Evidence. To test whether participants who had received a prior probability probe ($N = 111$) misaggregated the probative value of the DNA test results with their prior probability of guilt estimates, we used Bayes' theorem to calculate the normatively appropriate posterior probability of guilt estimates and compared the result with participants' actual posterior probability of guilt estimates. Prior probability of guilt estimates ($M = .60$, $SD = .19$) did not differ across conditions, $F(3, 107) = 1.00$, *ns*—an expected finding given that participants had not yet been exposed to any between-subjects experimental manipulations.

We crossed participants' actual posterior probability of guilt estimates and the Bayesian normative estimates with the type of probabilistic information to which they were exposed using a 2 (actual versus Bayesian posterior probability of guilt estimate) by 4 (SE, TSE, TSECI, or SCE probabilistic information condition) mixed model ANOVA. Consistent with Hypothesis 8, participants' actual posterior probability of guilt judgments ($M = .85$, $SD = .17$) were significantly lower than Bayesian normative estimates ($M = .99$, $SD = .02$), $F(1, 107) = 83.04$, $p < .05$, indicating that participants misaggregated the probabilistic evidence with their prior probability of guilt estimates. No other significant effects emerged.

Hypothesis 9: Personal Estimates. Did participants accept the expert's estimates of RMP and LE at face value? Of participants who correctly recalled the probability estimates the criminalist had mentioned, the modal personal estimate of RMP and LE was the same as the expert's estimate.[12] However, the rate at which participants endorsed the expert's numbers ranged from a low of 19% to a high of 46%, and thus in no instance did even a simple majority of participants accept the criminalist's estimate. When considering the extremely small (i.e., 1 in a billion) probability estimate, participants' personal estimates of RMP and LE were nearly always equal to or greater than the expert-provided estimate. However, this was not the case when participants considered the larger (i.e., 2 in 100) estimate: Fifty-six percent of participants who evaluated an LE of 2 in 100 indicated that their personal estimate of LE was equal to or greater than the expert-provided estimate, as compared to only 39% of participants who evaluated an RMP of 2 in 100.

What were participants' expectancies of the probability that the match report was due to random chance or laboratory error when the criminalist did not provide such information? The mean uninformed LE estimate of participants who received only RMP information ($N = 29$) was just under 1 in 15 ($M = .07$, median $= .001$, $SD = .14$), while the mean uninformed RMP estimate of participants who received only LE information ($N = 31$) was approximately 1 in 333 ($M = .003$, median $=$ $.000001$, $SD = .01$).[13] Note that this finding is consistent with Hypothesis 5: Par-

[12]There was one minor exception to this finding. In the TSECI (RMP = 2 in 100) condition, the expert-provided RMP estimate was chosen 4% less frequently than the 1 in 1,000 modal RMP estimate jurors accepted.

ticipants' mean uninformed estimates of LE were about 20 times larger than their uninformed estimates of RMP for the trimmed distributions (and about 4 times larger for the untrimmed distributions). The t tests performed on both the trimmed and untrimmed distributions revealed that this difference was significant for the trimmed data, $t(56) = 2.49, p < .05$, although not for the untrimmed data, $t(56) = 1.59$, *ns*.

Participants who were given a single combined estimate of 2 in 100 that the match report occurred due to either random chance or human error in the DNA laboratory ($N = 28$) had a mean LE expectancy of about 1 in 17 ($M = .06$, median $= .02$, $SD = .13$). They had a mean RMP expectancy of 1 in 25 ($M = .04$, median $= .01$, $SD = .08$).

What were participants' expectancies of the probability that the match report was due to intentional tampering? The mean participant estimate of the probability that the match report occurred due to intentional tampering by the police, prosecutor, or criminalist ($N = 219$) was about 1 in 50 ($M = .02$, median $= .0002$, $SD = .08$). The size of this estimate did not vary across experimental conditions.

To test whether incorporating participants' personal estimates and expectancies into Bayes' theorem would attenuate the gap between their posterior probability of guilt estimates and Bayesian norms, we recomputed Bayesian posterior probability of guilt estimates by summing participants' own personal estimates of RMP, LE, and intentional tampering and subtracting the corresponding probability of their joint occurrence. We entered actual posterior probability of guilt estimates, Bayesian posterior probability of guilt estimates (incorporating only the expert-provided estimates of RMP and LE), and the recomputed Bayesian posterior probability of guilt estimates into a 3 (actual, Bayes, or Bayes-recomputed posterior probability of guilt estimate) by 4 (SE, TSE, TSECI, or SCE probabilistic information condition) mixed model ANOVA. A main effect emerged for type of posterior probability of guilt estimate, $F(2, 104) = 41.78, p < .05$. Consistent with Hypothesis 9, follow-up tests indicated that participant expectancies did attenuate the gap between actual and Bayesian posterior probability of guilt estimates. The Bayesian recomputed estimates were significantly lower ($M = .94$, $SD = .15$) than the original Bayesian estimates ($M = .99$, $SD = .02$), $F(1, 105) = 7.47, p < .05$, but were still significantly higher than participants' actual posterior probability of guilt judgments ($M = .84$, $SD = .17$), $F(1, 105) = 27.09, p < .05$. These data indicate that even though participants' expectancies and personal estimates of RMP, LE, and intentional tampering account for some of the gap between actual posterior probability of guilt estimates and Bayesian norms, decision makers still appear to underweight probabilistic DNA evidence.

Summary of Subjective Probability Data. These data cast further doubt on the viability of the vividness hypothesis as an explanation of the misperception error. This is not terribly surprising, however, given the various, and mostly unsuccessful,

[13]The statistics relevant to Hypothesis 9 are based on a distribution of personal estimates in which extremely high scores (above the 97th percentile) were trimmed by assigning them the value of the estimate given by the participant at the 97th percentile. Analyses of the untrimmed data yielded substantially similar results.

post-1980 attempts to find empirical support for the vividness hypothesis in the social psychology research lab. In a meta-analysis of 25 studies, Taylor and Thompson (1982) found almost no support for the vividness hypothesis. More recent research by Collins, Taylor, Wood, and Thompson (1988) revealed that although vivid information was more memorable than pallid information, it was not more persuasive and did not result in heightened attitude change. Thus it is worthwhile to examine other possible explanations for why people tend to overweight extremely small probabilities in their judgments, such as the averaging strategy hypothesis, which did receive some support from our combination strategy data.

These data also lend support to the expectancy-based description of how people evaluate DNA evidence. Participants' mean uninformed estimate of LE was approximately 1 in 15, which is about 3 times larger than the expert-provided LE of 2 in 100 and about 22 times larger than participants' mean uninformed estimate of RMP (which was approximately 1 in 333). This finding is consistent with participants' tendency to convict less often when their LE expectancies went unchecked than when their RMP expectancies went unchecked.

Finally, consistent with prior Bayesian updating studies and the review by Saks and Kidd (1980), participants appeared to underweight the probabilistic evidence by misaggregating their prior probability of guilt estimates and the probative value of the DNA test results. However, this finding was partially attenuated when participants' expectancies were modeled using Bayes' theorem, suggesting that expectancies also play a role in the formation of subjective probabilities.

DISCUSSION

We began this paper by proposing that reactions to DNA evidence can be described both in terms of errors and expectancies, and we designed an experiment that allowed us to investigate the role played by each in ways that prior studies did not permit. The results of this study provide evidence that evaluations of DNA evidence are influenced by both systematic errors and lay expectancies about the sources of a DNA match report.

Implications of Errors

When examined from an errors perspective, the results of this study suggest that jurors generally underweight probabilistic evidence in their decisions (when compared to Bayesian norms) via the misaggregation error. However, when jurors receive two separate probability estimates they appear to overweight the extremely small probability by misperceiving how the two estimates should be combined. The misperception error appears to persist even when jurors are provided with a simple combination instruction.

Should Jurors Continue to Receive Separate RMP and LE Estimates?

Given decision makers' apparent inability to combine separate RMP and LE estimates appropriately, Koehler (1997; Koehler et al., 1995) argued that jurors should be presented with an aggregated estimate of the probability that the match occurred due to either random chance or human error in the lab, perhaps in the form of a likelihood ratio. In this way jurors will not receive (and correspondingly will not be unduly swayed by) extremely small, but essentially irrelevant, RMP estimates. In contrast, the National Research Council Report (1996) rejected the notion of presenting jurors with only an aggregated estimate of the probability that a match occurred due to either random chance or laboratory error on a variety of grounds, including the argument that aggregated estimates deprive fact finders of valuable information (i.e., fact finders cannot separately evaluate the possibility of a coincidental match and the possibility of a match due to laboratory error).

We prefer the position of the National Research Council Report (1996). Jurors, as fact finders, may need to know the disaggregated elements that influence the aggregated estimate as well as how they were combined in order to evaluate the DNA test results in the context of their background beliefs and the other evidence introduced at trial. For instance, if jurors do not accept the expert-provided probability associated with the aggregated estimate (as many participants in this study did not), they have neither the disaggregated estimate of LE to guide their decisions, nor the knowledge that whatever value they attach to the LE estimate will likely render irrelevant the extremely small value they attach to the RMP. This can be problematic for a variety of reasons. Consider the juror who is wary of the competence of lab technicians at the lab where the DNA was analyzed and decides that there is a 1 in 10 chance that the match occurred due to LE. Will this juror understand that the aggregated estimate should now also be 1 in 10? Or will this juror average the two estimates together (or engage in some other combination strategy) and arrive an incorrectly aggregated estimate that is smaller than 1 in 10? In this case the juror who receives the aggregated estimate is no better off than the juror who receives the separate component estimates.

Can Jurors Learn to Combine Separate RMP and LE Estimates Appropriately?

Attempts to improve jurors' use of probabilistic evidence have not met with much success. Simple explanations of how to apply Bayes' theorem (e.g., Faigman & Baglioni, 1988; Smith et al., 1996), presenting probabilistic information in graphical form (e.g., Goodman, 1992), and providing simple instructions about how to combine separate probability estimates (e.g., this study) have not helped decision makers become better consumers of probabilistic evidence. A question to ask is: Did these instructions increase decision makers' awareness of the logical errors they were committing and correspondingly increase their ability and/or motivation to correct these errors?

With regard to ability, results of the Faigman and Baglioni (1988) study revealed that participants' understanding of Bayes' theorem was still minimal even after instruction. Similarly, participants in this study who received combination instructions had less than perfect recall for those instructions. Although these results may indicate that people *cannot* evaluate probabilistic evidence appropriately, they may instead demonstrate the failure of researchers in this area to utilize effective statistical training techniques that have been identified in recent research (e.g., Fong, Lurigio, & Stalans, 1990; Hanita, Gavanski, & Fazio, 1997).

With regard to motivation, the simple combination instruction employed in this study may have been more effective if it had included an explanation of why it was important to combine the extremely small and much larger probability appropriately, rather than a simple admonition to do so. Diamond and Casper (1992) found that jurors who were provided with an explanation of why they were not to consider the judge's duty to treble the amount of punitive damages they wished to award a plaintiff in a civil case gave more appropriate awards than jurors who were simply admonished not to consider this information. Future research should explore whether these ability- and motivation-based strategies can improve jurors' use of probabilistic evidence.

Implications of Expectancies

When examined from an expectancies perspective, the results of this study suggest that we need to be more complete in our modeling of the Bayesian updating process: Bayesian norms are likely to be artificially inflated if expectancies (such as intentional tampering in this case) are not included in the model. The results of this study also suggest that jurors may not infer that DNA test results are error-free when they do not receive an LE estimate. Thus the "RMP-only" condition in the original Koehler et al. (1995) study might more appropriately be labeled the "RMP-plus-LE-expectancy" condition, as people appear to fill in the missing LE information with their own beliefs. Finally, people do not evaluate RMP and LE information in the same way even though their implication is identical (i.e., a match is declared even though the defendant is not the true source of the crime scene sample). Consequently, mathematically equivalent DNA test results are not always seen as psychologically equivalent—a finding that both complicates and enriches our understanding of how jurors evaluate probabilistic evidence.

What Other Expectancies Can Influence Evaluations of DNA Evidence?

Probabilities do not merely appear out of thin air, even though it may seem this way to participants in a mock trial simulation where they simply exist in black and white on a page. Jurors may have beliefs about the credibility of the experts who calculate such numbers, the criminalists who conduct such laboratory tests, and the trustworthiness of the police officers who collect the lab samples. Jurors may also have expectancies about sources of a reported match other than RMP, LE, or intentional tampering. When asked, participants in this study were able to provide a variety of additional explanations of how there could have been a reported match even though the defendant was not the true perpetrator. These ranged from the possibility of an identical twin being the attacker to the belief that the defendant and victim had consensual sex. To the extent that participant expectancies are not experimentally manipulated—or at least assessed and accounted for, as they were in this study—our ability to model the probabilistic inference process is weakened.

What About Inaccurate Expectancies?

Although jurors are permitted to evaluate evidence in the context of their own background experiences, it should be noted that these beliefs are not necessarily accurate reflections of the real world. Expectancies can be shaped by things like the media, the community one lives in, and other idiosyncratic life experiences. Left unchecked, inaccurate expectancies may lead to legally undesirable outcomes. Once inaccurate expectancies are identified, the court can take action to reduce their influence by admitting relevant expert testimony or providing appropriate jury instructions.

Next-Stage Research to Assess the Generalizability of These Results

In light of the fact that we cannot reliably predict when a student sample is likely to provide an adequate model of actual juror behavior, Diamond (1997) cautions researchers against basing policy recommendations on this kind of "Stage One" research. For instance, in this study we may have systematically underestimated the tendency of actual jurors to commit mathematical errors because the undergraduate participants were better educated than members of a typical jury. In addition, undergraduates' expectancies may differ from those of actual jurors because expectancies are likely formed over time as a result of life experience—and undergraduates are generally younger and less experienced with "the real world" than actual jurors. Whether actual jurors are more or less trusting of DNA test results than undergraduates is unknown. The effects of differences between undergraduates and actual jurors need to be tested in future research. The results here indicate, however, that even educated decision-makers may be susceptible to errors when evaluating complex scientific evidence, that simple instructions are unlikely to correct these errors, and that decision-maker expectancies may influence judgments.

Finally, the participants in this study responded individually—unlike jurors who have the opportunity to deliberate. Although we cannot tell whether deliberations would have reduced the influence of errors and expectancies on verdicts, the evidence on the ability of deliberations to correct errors is mixed (e.g., Diamond & Levi, 1996; Ellsworth, 1989; Hastie, Penrod, & Pennington, 1983). Where error rates are low, deliberations appear to help, but where error rates are high, discussions simply permit the exchange of misinformation (Diamond & Levi, 1996). In this study, error rates were high with regard to how participants combined separately presented RMP and LE estimates. Thus, deliberations would have been unlikely to help in this case unless at least one member of the deliberating group knew how to combine the separate estimates correctly and could convince the rest of the group that summing was the correct approach.

Are the Effects of Errors and Expectancies Observed here Limited to Laypersons?

What little research has been done to compare judge and jury decision making suggests that judges, too, can be influenced by errors and expectancies. For instance, Kalven and Zeisel (1966) found that the rate of disagreement between the judge and jury verdicts was not attributable to an inability of the jury to understand the evidence—the rate was the same whether the judge characterized it as "easy" or "difficult" to understand. In a similar vein, Wells (1992) found that judges and jury-eligible undergraduates were equally reluctant to find in favor of a plaintiff in a civil case when presented with naked statistical evidence. Furthermore, a substantial proportion of judges who chose not to apply statistical evidence gave weak or flawed reasons for discounting it. Finally, Landsman and Rakos (1994) found that judges and jurors who were exposed to information that had been ruled inadmissible were influenced by that information to a similar extent. Cognitive limitations and leaps in inference are not the exclusive province of the jury.

CONCLUSION

In addition to shedding empirical light on the specific hypotheses of interest, the results of this study contribute to our understanding of how legal decision makers evaluate complex scientific evidence more generally. For example, consider the new policy implemented by the FBI that when the probability of a random match exceeds 1 in 260 billion, the expert witness will not present a quantitative RMP estimate, but will instead tell the jury that the identity of the source has been conclusively established (Tribune News Services, 1997). How will this change in policy affect jury decision making? Although the results of this study do not answer this question directly, they do suggest a number of relevant empirical questions that should be asked: Will the presentation of a quantitative LE estimate with the qualitative statement of identity focus juror attention on the LE? Will jurors be more or less impressed by a statement of identity than by an extremely small RMP estimate? The actual implications of this policy await future research.

To summarize, although jurors are often criticized as incapable of evaluating complex scientific evidence, ironically many of these criticisms are not themselves scientifically well grounded. The results of this study qualify the image that jurors are poor consumers of scientific evidence and point to a more complex picture of juror decision making as shaped by both errors and expectancies.

ACKNOWLEDGMENTS

This research is based on the first author's master's thesis and was partially funded by an American Psychology-Law Society Grant-in-Aid to the first author. An earlier version of this paper was presented at the Annual Meeting of the Law and Society Association, Glasgow, Scotland, July 1996. We would like to thank

Jonathan Koehler, Richard Lempert, Richard Wiener, and three anonymous reviewers for their helpful comments.

REFERENCES

Bernstein, D. E. (1996). Junk science in the United States and the Commonwealth. *Yale Journal of International Law, 21,* 123–182.
Collins, R., Taylor, S., Wood, J., & Thompson, S. (1988). The vividness effect: Elusive or illusory? *Journal of Experimental Social Psychology, 24,* 1–18.
Commonwealth v. Curnin, 565 N.E. 2d 440, 441 (Mass. 1991).
Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by juries and exonerated by science: Case studies in the use of DNA evidence to establish innocence at trial.* National Institute of Justice Research Report No. NCJ 161258.
Dellios, H. (1995, September 18). U.S. justice system hit by Simpson trial fallout. *The Chicago Tribune,* pp. A1, A10.
Diamond, S. S. (1997). Illuminations and shadows from jury simulations. *Law and Human Behavior, 21,* 561–571.
Diamond, S. S., & Casper, J. D. (1992). Blindfolding the jury to verdict consequences: Damages, experts, and the civil jury. *Law & Society Review, 26,* 513–563.
Diamond, S. S., & Levi, J. N. (1996). Improving instructions on death by revising and testing jury instructions. *Judicature, 79,* 224–232.
DuCharme, W. (1970). Response bias explanation of conservative human inference. *Journal of Experimental Psychology, 85,* 66–74.
Ellsworth, P. C. (1989). Are twelve heads better than one? *Law and Contemporary Problems, 52,* 205–224.
Faigman, D., & Baglioni, A. (1988). Bayes' theorem in the trial process. *Law and Human Behavior, 12,* 1–17.
Fong, G. T., Lurigio, A. J., & Stalans, L. J. (1989). Improving probation decisions through statistical training. *Criminal Justice and Behavior, 17,* 370–388.
Funder, D. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.
Goodman J. (1992). Jurors' comprehension and assessment of probabilistic evidence. *American Journal of Trial Advocacy, 16,* 361–389.
Hanita, M., Gavanski, I., & Fazio, R. H. (1997). Influencing probability judgments by manipulating the accessibility of sample spaces. *Personality and Social Psychology Bulletin, 23,* 801–813.
Hastie, R., Penrod, S. D., & Pennington, N. (1983). *Inside the jury.* Cambridge, MA: Harvard University Press.
Hays, W. L. (1994). *Statistics (5th ed.).* Fort Worth, TX: Harcourt Brace.
Huber, P. W. (1991). *Galileo's revenge: Junk science in the courtroom.* New York: Basic Books.
Imwinkelried, E. (1982–1983). The standard for admitting scientific evidence: A critique from the perspective of juror psychology. *Villanova Law Review, 28,* 554–571.
Kalven, H., & Zeisel, H. (1966). *The American jury.* Chicago, IL: University of Chicago Press.
Koehler, J. J. (1993). Error and exaggeration in the presentation of DNA evidence at trial. *Jurimetrics Journal, 34,* 21–39.
Koehler, J. J. (1997). Why DNA likelihood ratios should account for error (even when a National Research Council Report says they should not). *Jurimetrics Journal, 37,* 425–437.
Koehler, J. J., Chia, A., & Lindsay, S. (1995). The random match probability in DNA evidence: Irrelevant and prejudicial? *Jurimetrics Journal, 35,* 33–58.
Landsman, S., & Rakos, R. F. (1994). A preliminary inquiry into the effect of potentially biasing information on judges and jurors in civil litigation. *Behavioral Sciences and the Law, 12,* 113–126.
Lempert, R. (1993). DNA, science, and the law: Two cheers for the ceiling principle. *Jurimetrics Journal, 34,* 41–57.
MacCoun, R., & Kerr, N. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology, 54,* 21–33.
National Research Council Report. (1992). *DNA technology in forensic science.* Washington, DC: National Academy Press.
National Research Council Report. (1996). *Report on the evaluation of forensic DNA evidence.* Washington, DC: National Academy Press.

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

People v. Collins, 68 Cal. 2d 319 (1968).

Peterson, J. L., Ryan, J. P., Houlden, P. J., & Mihajlovic, S. (1987). The uses and effects of forensic science in the adjudication of felony cases. *Journal of Forensic Sciences, 32,* 1730–1753.

Sanders, J. (1993). From science to evidence: The testimony on causation in the Bendectin cases. *Stanford Law Review, 46,* 1–86.

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

Schklar, J. (1996, March). *Naive beliefs about science and reactions to DNA evidence.* Paper presented at the 1996 American Psychology—Law Society Biennial Conference in Hilton Head, South Carolina.

Schklar, J. (1996, July). *Juror reactions to DNA evidence: Misperception and misaggregation.* Paper presented at the 1996 Law and Society Conference in Glasgow, Scotland.

Smith, B. C., Penrod, S. D., Otto, A. L., & Park, R. C. (1996). Jurors' use of probabilistic evidence. *Law and Human Behavior, 20,* 49–82.

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

Thompson, W. C. (1993). Evaluating the admissibility of new genetic identification tests: Lessons from the "DNA war." *Journal of Criminal Law and Criminology, 84,* 22–104.

Thompson, W. C., & Schumann, E. (1987). Interpretation of statistical evidence in criminal trials. *Law and Human Behavior, 11,* 167–187.

Tribe, L. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review, 84,* 1329–1376.

Tribune News Services. (1997, November 13). DNA analysis now can leave no doubt, FBI reports. *Chicago Tribune,* p. A15.

United States v Addison, 498 F. 2d 741 (1974).

Wells, G. L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62,* 739–752.

Wilson, T., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted thought influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.