Law and Human Behavior, Vol. 25, No. 5, October 2001 (© 2001)

When Are People Persuaded By DNA Match Statistics?[1]

Jonathan J. Koehler[2]

The way in which statistical DNA evidence is presented to legal decision makers can have a profound impact on the persuasiveness of that evidence. Evidence that is presented one way may convince most people that the suspect is almost certainly the source of DNA evidence recovered from a crime scene. However, when the evidence is presented another way, a sizable minority of people equally convinced that the suspect is almost certainly not *the source of the evidence. Three experiments are presented within the context of a theory ("exemplar cueing theory") for when people will find statistical match evidence to be more and less persuasive. The theory holds that the perceived probative value of statistical match evidence depends on the cognitive availability of coincidental match exemplars. When legal decision makers find it hard to imagine others who might match by chance, the evidence will seem compelling. When match exemplars are readily available, the evidence will seem less compelling. Experiments 1 and 2 show that DNA match statistics that target the individual suspect and that are framed as probabilities (i.e., "The probability that the suspect would match the blood drops if he were not their source is 0.1%") are more persuasive than mathematically equivalent presentations that target a broader reference group and that are framed as frequencies ("One in 1,000 people in Houston would also match the blood drops"). Experiment 3 shows that the observed effects are less likely to occur at extremely small incidence rates. Implications for the strategic use of presentation effects at trial are considered.*

Most people agree that forensic DNA analysis represents an important theoretical advance for the criminal justice system. Indeed, when DNA evidence arrived in U.S. courtrooms more than a decade ago, one court excitedly referred to it as "the single greatest advance in the 'search for truth,' and the goal of convicting the guilty and

[1]Parts of this research were presented at the 1998 American Association for the Advancement of Science meeting in Washington, DC, the 1997 American Psychological Association meeting in Chicago, the 1998 American Psychology-Law Society meeting in Redondo Beach, the 1999 European Science Foundation meeting in Milan, the 1999 International Conference on Forensic Statistics in Raleigh, and the 1998 Judgment and Decision Making Society meeting in Dallas.

[2]McCombs School of Business and School of Law, The University of Texas at Austin, CBA 5.202, Austin, Texas 78712-1175; e-mail: koehler@mail.utexas.edu.

0147-7307/01/1000-0493$19.50/1 © 2001 American Psychology-Law Society/Division 41 of the American Psychology Association

acquitting the innocent, since the advent of cross-examination" (*New York v. Wesley*, 1988, p. 644).

The value of DNA analyses for criminal justice purposes lies in its theoretical power to exclude large proportions of the population as potential contributors of genetic material (e.g., blood, semen, hair) that are recovered from violent crime scenes. Thus, when a suspect's DNA matches the DNA recovered from a crime scene sample, a prosecutor can argue that the suspect is the likely source of that sample and therefore was present at the crime scene. In most cases, the probative value of a DNA match is conveyed to triers of fact via statistics that describe how common the characteristics of the DNA profile are in a given population (Aitken, 1995; Evett & Weir, 1998; Fienberg, 1989; Kaye, 1993; Kaye & Sensabaugh, 2000; National Research Council, 1992; Robertson & Vignaux, 1995).

As documented in the first National Research Council report on DNA evidence, there was scientific controversy in the early 1990s about how to estimate profile frequencies (National Research Council, 1992). There is less disagreement now (Lander & Budowle, 1994; see Mueller, 1993, for a different view). However, a second National Research Council Report raised another disturbing issue. In a three-paragraph section entitled "the importance of behavioral research," it concluded that "we know very little about how laypersons *respond* to DNA evidence" (National Research Council, 1996, p. 203, italics added). This conclusion was correct in the narrow sense that almost no research had been undertaken on how different factors affect the weight that people assign to DNA statistics. However, as the report acknowledged, a large body of research suggests that people have poor intuitions when it comes to statistical reasoning in general (see e.g., Arkes, Connolly, & Hammond, 2000; Hogarth, 1991; Kahneman, Slovic, & Tversky, 1982; Saks & Kidd, 1980/1981).

Poor statistical reasoning among laypeople has also been observed in legal decision-making studies. Research with non-DNA statistical evidence showed that, in general, people attach less weight to the statistical evidence than would seem appropriate (Faigman & Baglioni, 1988; Goodman, 1986, 1992; Smith, Penrod, Otto, & Park, 1996; Taroni & Aitken, 1998a; Thompson & Schumann, 1987), and are insensitive to variations in the diagnosticity of the statistical evidence (Faigman & Baglioni, 1988; Goodman, 1986, 1992; Thompson, Britton, & Schumann, 1990; Thompson & Schumann, 1987; for early reviews see Kaye & Koehler, 1991; Thompson, 1989).

Research with DNA statistics—where the statistical evidence tends to be more probative—tells a similar tale. Koehler, Chia, and Lindsey (1995) showed that mock jurors had trouble aggregating a 1 in 1 billion DNA match statistic with laboratory error rate statistics. Schklar and Diamond (1999) replicated this finding and showed that, consistent with earlier studies on non-DNA statistical evidence, jurors underestimated the probative value of DNA evidence relative to Bayesian norms. Taroni and Aitken (1998b) found that the DNA statistics were undervalued even among forensic scientists and forensic science students. Rowe (1997) showed that jurors do a better job aggregating a moderately probative DNA match statistic (.01) with nonscientific evidence when presented with a Bayesian-like judicial instruction. However, Smith et al. (1996) did not find that Bayesian instruction was helpful in a blood-and-enzyme match case, using slightly less probative statistical evidence. Koehler (1996) found that laypeople tend to be more impressed with DNA statistics when they are presented as likelihood ratios rather than as frequencies. This may have occurred because untrained jurors mistakenly equated the likelihood ratios with posterior odds ratios. Interestingly, Taroni and Aitken (1998b) observed a similar pattern in a sample of lawyers but the opposite pattern in a sample of scientists and science students.

Taken as a whole, the evidence to date suggests that laypeople are not intuitive Bayesians in cases involving DNA statistics, and they may not assess the probative value of a DNA match in clear and consistent ways. In light of the voluminous research on probabilistic reasoning shortcomings, these conclusions may not be surprising. However, they are useful because they motivate the question of whether jurors' use of DNA statistics can be influenced and predicted. This paper focuses less on examining *how well* jurors evaluate DNA evidence and more on the psychological question of *how* jurors assess the value of DNA evidence. Toward this end, I heed the advice of Smith et al. (1996, p. 79) and offer a theory for predicting when legal decision makers will be more and less impressed by DNA match statistics. Following a discussion of the broad theory and its significance for cases involving DNA evidence, three experiments with using a mock juror paradigm are presented. The experiments show how subtly different presentations of the identical DNA match statistics in a murder case affect (a) estimates of the chance that the suspect is the source of the matching DNA evidence, (b) estimates of the chance that the suspect is guilty, and (c) verdicts.

EXEMPLAR CUEING THEORY

When a DNA match is reported, the strength of that match is typically conveyed via the "random match probability" (RMP; National Research Council, 1996, p. 31). The RMP identifies the frequency of a DNA profile in a reference population. To the extent that the RMP is low (e.g., one in many thousand, million, or billion), the chance that the match between a recovered crime scene sample and a suspect is unlikely to be purely coincidental.

In order to predict when legal decision makers will be more and less persuaded by the RMP, this research draws on exemplar cueing theory (Koehler, in press; Koehler & Macchi, 2001). This theory is based on the assumption that people evaluate the significance of low probability events, in part, as a function of the ease with which they are able to imagine examples of the events. Where the probabilistic information in question identifies how common a matching profile is in some reference population, exemplar cueing theory holds that people judge the probative value of the reported match by the ease with which they can imagine examples of others who would also match. I refer to these as "coincidental match exemplars" or "match exemplars." When people find it hard to imagine match exemplars, the evidence will be treated as compelling proof of identity (i.e., proof that the matching suspect is the source of the recovered DNA evidence). But when match exemplars are easier to imagine, the evidence will seem less compelling. I will argue that variables related to the presentation of mathematically identical DNA evidence affects the ease with which coincidental match exemplars are called to mind (or otherwise become salient).

Related Theory

Exemplar cueing theory has its roots in the availability heuristic (Tversky & Kahneman, 1973). Availability holds that people judge the frequency or probability of an event by the ease with which exemplars come to mind. Unlike availability, exemplar cueing theory is not a heuristic that people use to estimate a frequency or a probability. Instead, it is a theory about how people determine the subjective weight to assign to the hypothesis that the matching suspect is the source of the recovered evidence.

Exemplar cueing theory is also related to recent research by Slovic, Monahan, and MacGregor (2000, Study 3). These researchers provided clinicians with written descriptions of psychiatric patients and asked the clinicians to judge (among other things) the risk that the patients would harm someone if discharged from a mental health facility (see also Slovic & Monahan, 1995). Slovic et al. (2000) found that the clinicians' judgments varied depending on how base rate information about harm caused by similar other patients was presented. When the base rate was provided as a relative frequency (e.g., "of every 100 patients similar to Mr. Jones, 10 are estimated to commit an act of violence"), the clinicians were less likely to recommend an immediate patient discharge than when the base rate was provided as a probability (e.g., "patients similar to Mr. Jones are estimated to have a 10% probability of committing an act of violence"). The authors explained this effect through reference to an affect heuristic (Finucane, Alhakami, Slovic, & Johnson, 2000): clinicians who received the risk information as a relative frequency were relatively more likely to have "frightening images" of acts of violence (Slovic et al., 2000).

The Slovic et al. (2000) data and the "frightening images" explanation may be regarded as a special case of exemplar cueing. As discussed later, exemplar cueing captures the idea that there are ways to communicate probabilistic data that encourage decision makers to think beyond the narrow confines of the target case. By presenting clinicians with a hypothetical set of 10 violent patients out of a set of 100 patients, violent exemplars were readily available in the minds of the clinicians. These examples may have affected judgments about the focal case. No such exemplars were cognitively available when the risk base rate was presented as a probability. A significant difference between the frightening images explanation Slovic et al. (2000) offered for their data and the exemplar cueing theory explanation offered here is that the latter does not require an affective explanatory component. Instead, exemplar effects may operate by providing a vivid reminder to decision makers that low probability events (including those that are devoid of the strong affective imagery), which may be represented in mind as "zero" (Stone, Yates, & Parker, 1994), are actually associated with one or more "real" outcomes.

The exemplar cueing framework is also consistent with research that identifies two ways in which the human mind reasons with statistical information (Sloman, 1996). Under a "distributional" (Hansen & Helgeson, 1996; Kahneman & Tversky, 1982; Klar, Medding, & Sarel, 1996; Reeves & Lockhart, 1993), "extensional" (Tversky & Kahneman, 1983), or "rational" (Denes-Raj & Epstein, 1994; Epstein, Lipson, Holstein, & Huh, 1992; Kirkpatrick & Epstein, 1992; Miller, Turnbull, & McFarland, 1989) way of reasoning, people construct reference classes, attend to base rates, and otherwise reason in accordance with the rules of probability. Kahneman and Lovallo (1993) refer to this approach as the "outside view" because it requires thinking about target events from the outside as a member of a larger class of events. In contrast, under a "singular," "nonextensional," or "experiential" way of thinking, people tend to think about target events in isolation. Kahneman and Lovallo (1993) refer to this approach as the "inside view" because it requires thinking about target events from the inside without reference to related cases.

The distinction made by these two-system models and the inside/outside dichotomy that Kahneman and Lovallo (1993) invoke is central to exemplar cueing theory as well. An outside perspective cues coincidental match exemplars because it provides both event probabilities and a reference class within which to search for and generate exemplars. In contrast, an inside perspective does not offer a readily available reference class and therefore does not cue coincidental match exemplars.

Target and Frame

One possible way to cue coincidental match exemplars in the minds of legal decision makers is to adjust the "target" of the DNA match away from the focal suspect and onto a larger reference population. For example, jurors may be told that, although the suspect's DNA matches the recovered DNA evidence, DNA profiles of 1 in every 1,000,000 people in a large reference population (e.g., the city in which the crime occurred) would also match. This approach, which will be referred to as "multitarget," is designed to focus decision makers on the possibility that there are others who match but who are not the source of the DNA evidence. Jurors who are encouraged to think about multiple targets may be less impressed with and persuaded by the statistical DNA evidence. In contrast, when the target of the DNA match remains on the focal suspect (e.g., "the chance that *the suspect* would match by coincidence if he were not the source is 1 in 1,000,000"), jurors are less likely to think about others who might match. This will be referred to as a "single-target" approach.

A second factor that may influence whether jurors think about coincidental match exemplars is whether the DNA statistic is "framed" as a frequency or as a probability. A DNA statistic could be framed as "1 in 1,000,000" ("frequency frame") or as ".000001" ("probability frame"). The two frames are mathematically—but not psychologically—identical. A growing body of research shows that people reason differently with frequencies than with probabilities (Cosmides & Tooby, 1996; Gigerenzer, 1998; Gigerenzer & Hoffrage, 1995; Hansen & Helgeson, 1996; Harries & Harvey, 2000; Reeves & Lockhart, 1993; Slovic et al., 2000; Tiegen, Brun, & Frydenlund, 1999; Tversky & Kahneman, 1983). Using, once again, terminology from Kahneman and Lovallo (1993), frequency frames induce people to adopt a broad, "outside" view in which instant cases are thought of from the outside as but one member of a larger class of events. Probability frames induce people to adopt a narrower, "inside" view in which instant cases are contemplated in isolation.

Applying this idea to the DNA context, frequency framings of DNA statistics may induce jurors to cast a broad net when thinking about the DNA evidence. This, in turn, may encourage exemplar thinking (i.e., thoughts of others who might match by coincidence). In contrast, probability framings may keep jurors' attention narrowly focused on the suspect. This would discourage exemplar thinking. If true, then we would expect jurors to be more persuaded by DNA statistics that appear as probabilities than those that appear as frequencies.

To summarize, exemplar cueing theory predicts that the way in which statistical DNA evidence is presented influences the persuasiveness of that evidence. If the DNA statistic focuses on the suspect (single-target) or is framed as a probability (probability frame), then the evidence will be viewed as relatively impressive and strong proof of identity. If the presentation promotes thoughts about others who might also match—either through the introduction of a broad target class (multi-target) or the use of statistical frequencies (frequency frame)—the DNA evidence will be viewed as less impressive and will not be persuasive proof of identity. Three experiments were conducted to test these predictions.

EXPERIMENT 1: CLINTON–LEWINSKY

In late January 1998, the United States was abuzz with a rumor that President Clinton had an affair with former White House intern Monica Lewinsky. It was further rumored (though not substantiated until the following summer) that Ms. Lewinsky owned a dress that contained a semen stain that matched Mr. Clinton's DNA profile. In order to test whether there might be a relationship between ease of exemplar generation (as influenced by the presentation of the statistical DNA evidence) and the perceived strength and persuasiveness of the evidence, I conducted the following experiment.

Method

Participants

The participants were 72 jury-eligible undergraduate students (55% female) at the University of Texas who were currently enrolled in an introductory business class. The experiment was conducted in a classroom setting. Jury-eligibility was inferred from participants' responses to the following questions: (1) Are you a U.S. citizen? (yes); (2) Have you ever been convicted of a felony? (no); and (3) Are you at least 18 years old? (yes). Data from 12 non-U.S. citizens and one convicted felon were excluded from the original sample ($n = 85$).

Materials and Procedure

All participants were provided with a written stimulus (two pages in length) that began with factually accurate background information regarding the Clinton–Lewinsky issue. Next, all participants were asked to assume that a dress worn by Ms. Lewinsky did, in fact, contain "some genetic material (i.e., semen) that matched the DNA of President Clinton" and that "a DNA expert reports that his tests could not rule out Mr. Clinton as a possible source of the recovered genetic material."

In addition, half of the participants (selected at random) read the following sentence: "The probability that Mr. Clinton would match the semen stain if he were not its source is 0.1%." This presentation of the DNA evidence employed a single-target (Mr. Clinton) and a probability frame (0.1%). As discussed earlier, such a presentation is not conducive to the generation of coincidental match exemplars. The other half of the participants read the following sentence: "1 in 1,000 people in Washington who are not the source would also match the semen stain." This presentation of the DNA evidence employed a multi-target (people in Washington) and a frequency frame (1 in 1,000). This presentation is conducive to exemplar generation. Next, all participants read that Mr. Clinton's attorney denied the relevance of the DNA evidence and suggested that the lack of corroborating eyewitnesses amounted to a weak case against his client.

On the second page, participants were asked to identify "the probability that Mr. Clinton was the source of the DNA recovered from the dress," P(source). Participants also answered some demographic and jury-eligibility questions. I predicted that participants in the multi-target/frequency frame (m/f) group would assign lower P(source) values than those in the single-target/probability frame (s/p) group.

Results and Discussion

Participants' P(source) responses were highly variable, ranging from 0 to 100% on the P(source) question. This may have been the result of using a sex scandal vignette in a politically charged environment. Nevertheless, nonparametric analyses revealed that s/p participants thought that it was more likely that Mr. Clinton was the source of the DNA than m/f participants (s/p $Mdn = 82.5\%$, m/f $Mdn = 60.5\%$, $z = 1.96$, $p = .05$).[3,4]

One interesting way of examining the results is in terms of the proportion of participants who were certain or nearly certain that President Clinton was the source of the stain on the dress under the circumstances provided. The data showed that the proportion of participants who were *at least 99% certain* that President Clinton was the source of the genetic material dropped from 28% in the s/p condition to 8% in the m/f condition, χ^2 $(1, n = 72) = 4.60$, $p = .032$.

Of course, there are many variables at work in a loaded context such as this. Politics, the influence of the news media, prior beliefs about the credibility of Clinton and Lewinsky, prior beliefs about the relationship between Clinton and Lewinsky, etc. These extraneous effects may mask some of the information presentation effects of interest here. Nevertheless, the results indicated that the m/f presentation of DNA evidence may be less persuasive than the s/p presentation even in a politically charged environment.

[3] All p values are two-tailed.

[4] If the P(source) medians seem low for a case in which there is a reported DNA match, it may help to recall that this experiment was conducted (a) when President Clinton's approval rating was at an all-time high (73%; CBS News, January 29, 1998), (b) When 40% of Americans did not believe there was an affair (ABC News, January 30, 1998), and (c) after several major news orgaizations (erroneously) reported that the FBI did not find evidence of semen on clothing taken from Ms. Lewinsky (CBS News, January 29, 1998; ABC News, January 30, 1998; see also Shepard, 1998, p. A6). See Experiment 2 for additional discussion of the seemingly low persuasive power of DNA match satistics in laboratory experiments.

EXPERIMENT 2: TARGET AND FRAME EFFECTS

Having established that the presentation of DNA evidence can affect judgment in a rich, real world context, we now turn to the laboratory in an attempt to disentangle the target and frame variables and to determine what happens when legal decision makers are exposed to both target and frame presentations. Thus, Experiment 2 examines two questions. First, are DNA presentation effects due to target, frame, or both? The theory predicts main effects for both variables because each affects people's ability to generate match exemplars. Second, will the effects of target and frame persist when decision makers are presented with both target and frame perspectives? This question is important because, in an actual trial, the statistical evidence is likely to be referred to more than once and in different ways. For example, evidence that is presented in s/p form by the prosecution may be recharacterized by the defense in m/f form. How will "dual presentation" jurors respond?

One possibility is that dual presentation jurors will respond similarly to m/f jurors. This is because both types of jurors are exposed to a presentation form that induces them to imagine others who might match. By this reasoning, the dual presentation jurors and m/f jurors will be less impressed with the evidence than s/p jurors. A second possibility is that jurors will anchor on their initial impressions of evidence strength and pay less attention to the recharacterized statistical evidence. This is consistent with the anchoring and adjustment theory of decision making (Tversky & Kahneman, 1974) and the primacy effects predicted by the Belief Adjustment Model for end-of-sequence (i.e., single-judgment) decision-making tasks (Hogarth & Einhorn, 1992). Experiment 2 explores target, frame, and dual presentation effects on participants' judgments in a hypothetical criminal case that includes DNA evidence.

Method

Participants

The participants were 227 jury-eligible undergraduate business students (48% female) at the University of Texas. Students participated for extra credit in an introductory business law class. The setting and determination of jury eligibility were the same as in Experiment 1. Data from 15 non-U.S. citizens and one convicted felon were excluded from the original sample ($n = 243$).

Materials and Procedure

Participants were assigned at random to one of the six conditions and provided with a three-page written stimulus. The stimulus described the facts of a hypothetical murder case in Houston that included a DNA match against the suspect/defendant. A sample stimulus is provided in the Appendix.

The stimuli were identical across all six conditions with the exception of a single sentence that described the DNA match statistic. The DNA match statistic, which was fixed at 0.001 (1 in 1,000), was described in one of four mathematically equivalent ways to participants in Conditions 1–4. The four descriptions differed in terms of target (single = s, multi = m) and frame (probability = p, frequency = f) as follows: (1) s/p: *The probability that the suspect would match the blood drops if he were not their source is 0.1%*; (2) s/f: *The frequency with which the suspect would match the blood drops if he were not their source is 1 in 1,000*; (3) m/p: *0.1% of the people in Houston would also match the blood drops*; and (4) m/f: *One in 1,000 people in Houston would also match the blood drops.*

Participants in Conditions 5 and 6 received evidence in both s/p and m/f form. The s/p and m/f perspectives were used because they represent the two most different perspectives according to exemplar cueing theory. Condition 5 received the s/p perspective first and Condition 6 received the m/f perspective first.

After studying the case facts and evidence, participants (1) estimated the probability that the defendant was the source of the DNA evidence, P(source), (2) estimated the probability that the defendant committed the crime, P(guilt), and (3) provided a verdict. Instructions were provided to give numerical probability estimates for P(source) and P(guilt), and to find the defendant guilty "only if the evidence convinces you 'beyond a reasonable doubt' that [the defendant] is guilty of this crime." A fourth question asked participants to estimate the number of innocent people in a city of 500,000 people (Austin, TX) who would give a DNA match with the recovered blood evidence. This question was designed to provide insight into whether exemplar-conducive DNA evidence presentations also facilitate *accurate* exemplar thinking.

The theory predicted main effects for target and frequency across the primary dependent measures. Specifically, it predicted that participants in the frequency frame (f) and multi-target (m) presentations would assign lower P(source) values, lower P(guilt) values, and would return fewer guilty verdicts than jurors in the probability frame (p) and single-target (s) conditions. The theory also predicted that data from jurors in the dual presentation conditions would be similar to data from participants in the m/f conditions. Order effects on the dual presentation conditions were predicted by an anchoring and adjustment theory but not by exemplar cueing theory.

Results

Data from the dual perspective conditions did not differ as a function of the order in which the s/p and m/f perspectives appeared for any of the dependent measures. Therefore, the data from the two dual presentation conditions (Conditions 5 and 6) were collapsed across this order variable and treated as a single experimental condition.

A MANOVA revealed significant differences across the five conditions on P(source) and P(guilt), $F(8, 440) = 2.90$, $p = .004$. As Table 1 shows, participants in the s/p condition gave the highest estimates for P(source) and P(guilt), those in the m/f condition gave the lowest estimates, and those in the dual perspective, s/f, and m/p conditions gave estimates that were in between. Tukey post hoc contrasts showed that the dual presentation estimates were significantly lower than the s/p estimates on P(source), dual $M = 49.5\%$, s/p $M = 6.7\%$, $p = .008$, and P(guilt), dual $M = 46.4\%$, s/p $M = 68.3\%$, $p = .037$. The dual presentation estimates were

Table 1. Mean Probability Estimates and Conviction Rates as a Function of Target and Frame (Exp. 2)

Condition	P(source)	P(guilt)	Verdict (% guilty)
Single-target (%)			
Probability frame	76.7	68.3	25.9
Frequency frame	56.8	48.5	24.0
Multi-target (%)			
Probability frame	54.7	56.9	24.1
Frequency frame	38.1	37.7	03.3
Dual perspectives (%)	49.5	46.4	21.5

Note: Cell sizes range from 25 to 30 in the first four conditions. Cell size was 113 for the dual perspectives condition.

not significantly different from the m/f estimates on either of these dependent measures. A logistic regression revealed that differences among the five conditions on the Verdict dependent measure were marginally significant, χ^2 (4, $n = 224$) = 8.46, $p =$ .076. The proportions in the far right column of Table 1 suggest that this marginal effect is driven by the relatively small proportion of guilty verdicts returned by participants in the m/f condition (3.3%). A post hoc difference in proportions analysis using a conservative standard error term indicates that the 3.3% m/f conviction rate is significantly different from the 23.0% conviction rate that was found across the other conditions ($z = 2.00$, $p = .046$).

A MANOVA conducted with the four single-perspective conditions (Conditions 1–4 only) revealed main effects for target, $F(2, 108) = 6.47$, $p = .002$, and frame, $F(2, 108) = 4.55$, $p = .013$, on P(source) and P(guilt). Participants in the single-target conditions assigned higher P(source) and higher P(guilt) estimates than jurors in the multi-target conditions, P(source): *M*s = 66.7 and 46.4%, respectively; P(guilt): *M*s = 58.4 and 47.3%, respectively. Similarly, participants in the probability frame conditions assigned higher P(source) and higher P(guilt) estimates than jurors in the frequency frame conditions, P(source): *M*s = 65.7 and 47.4%, respectively; P(guilt): *M*s = 62.6 and 43.1%, respectively. A test of proportions on verdicts did not detect significant main effects for target ($z = 1.29$, $p = .197$) or frame ($z = 1.69$, $p =$ .091), although the conviction rates for both variables were in the predicted direction (target: single $P = 25.0\%$, multi $P = 13.6\%$; frame: probability $P = 25.0\%$, frequency $P = 12.7\%$).

Which Method Best Promotes Understanding? As noted at the outset, the experiments reported here were not designed to determine which method of presenting the DNA evidence promotes good reasoning. However, the fourth question concerning the number of others in a city of 500,000 who would also match the DNA evidence (see the Appendix) provided indicated that jurors had a particularly difficult time thinking correctly about exemplars when the evidence was presented in s/p form.

Based on the .001 DNA incidence rate that the participants received, there would be about 500 matches in a population of 500,000 people. Because some jurors may have felt that the suspect was one of the 500 whereas others may not, answers of 499, 500, and 501 were counted as correct. Across all conditions, 58.1% of jurors produced the right answer. Holding aside jurors in the dual presentation condition (where the rate of correct responses was 64.9%), the remaining four conditions were analyzed to determine whether the target and frame variables affected jurors' ability to produce the correct answer. We observed a frame effect in which jurors who received the evidence as a frequency were more likely to get the answer correct than jurors who received the evidence as a probability, Ps = 60.7 and 42.1%, respectively, Wald χ^2 $(1, n = 113) = 3.87$, $p < .05$. A corresponding target effect was not found.

Discussion

Overall, the data from Experiment 2 support exemplar cueing theory. Participants who received the DNA match statistic in the multi-target and frequency frame conditions were less impressed by the statistic (in terms of P(source) and P(guilt judgments) than jurors who received the statistic in the single-target and probability frame conditions. This follows from the theoretical claim that statistics that have multi-targets and frequency frames are more conducive to coincidental match exemplar generation than statistics that have single targets and probability frames. Finally, our data suggest that people who receive evidence in the m/f format are less inclined to return a guilty verdict than people who are exposed to other forms of evidence presentation.

Providing jurors with both the m/f and s/p perspectives appeared to eliminate the verdict effect. However, based on their P(source) and P(guilt) responses, these dual perspective jurors still appeared to be less impressed with the statistical evidence than jurors who only received the s/p perspective. This is consistent with the view that when jurors are provided with the m/f frame, they are more likely to generate match exemplars, which in turn reduces the perceived strength of the evidence.

Finally, jurors who received the DNA evidence in frequency form were more likely to correctly identify the number of people in a medium-sized city who would also match the DNA profile. This result is consistent with research that indicates that people reason better with frequencies than probabilities (Cosmides & Tooby, 1996; Gigerenzer, 1998; Gigerenzer & Hoffrage, 1995; Harries & Harvey, 2000). It also provides *some* support for the position that jurors are less confused by frequency presentations of DNA evidence than probability presentations (Koehler, 1996). However, the normative issue here is more complicated because it is not clear whether jurors gave the DNA evidence the weight it deserved in any condition. Holding laboratory error rate considerations aside, the incidence rate provided to jurors here was roughly equivalent to a 1000:1 likelihood ratio on the P(source) question. Unless jurors assessed the prior probability that the matching suspect was the source of the sample to be very low (e.g., <10%), a full weighting of the DNA evidence probably should have produced higher posterior P(source) estimates for many jurors than was observed here.

Why Isn't The DNA Evidence More Persuasive? The data from Experiment 2 are interesting for reasons that go beyond the central mission of the study. Apparently, many people do not think that a person who matches a critical DNA sample is necessarily the source of that DNA or guilty of the crime that is associated with the sample. Two comments are relevant here. First, the evidence in the experiment was probably weaker than jurors expected. Not only was the DNA incidence rate of 1 in 1,000 less impressive than the 1 in millions and billions that people may expect from

DNA match statistics, but the case did not include much corroborating evidence against the defendant. I employed a relatively moderate evidentiary case to avoid ceiling effects in which jurors from all experimental groups were firmly convinced of the defendant's guilt.

Second, participants' reluctance to assign extremely high P(source) and P(guilt) values in cases built on statistical evidence is now a consistent finding in the literature (Koehler et al., 1995; Nance, 2000; Schklar & Diamond, 1999; Smith et al., 1996; Wells, 1992). This suggests that legal decision makers can only be so impressed with statistical evidence. As the RMP gets smaller, the theoretical probative value of the DNA evidence gets larger. So why are people unwilling to give the evidence the weight it would seem to deserve? One possibility is that they understand that the value of DNA evidence is limited by the possibility that the reported match is not, in fact, a match (Koehler et al., 1995; Lempert, 1991a,b). Some empirical support for this explanation exists. Schklar and Diamond (1999) reported that the average mock juror assumes that the chance of an erroneous DNA match call is about 1 in 15. Thus, participants in Experiment 2 may have valued the DNA evidence with the psychological equivalent of likelihood ratios that were closer to 15:1 than 1000:1. A second possibility is that people regard a DNA match to be a form of "naked statistical evidence" and are unwilling to assign extreme weights to such evidence without corroborating evidence (Kaye, 1993; Wells, 1992). A related possibility is that people are simply unwilling to assign extreme probabilities based case-summary information of the sort that is common to this and many other laboratory studies. Consequently, the subjective probabilities reported here may not reflect the levels of certitude that jurors reach in the actual court cases.

EXPERIMENT 3: SMALLER MATCH PROBABILITIES

Thus far, exemplar cueing theory has been tested in experiments that used match probabilities of 1 in 1,000. But DNA match probabilities that find their way into the courtroom are often as small as 1 in millions, billions, or trillions (Koehler, 1997). What does exemplar cueing theory say about the persuasive value of different presentations of match probabilities that are as small as 1 in 1 billion? Will target and frame matter when the incidence rates are so small? The theoretical answer is that it depends on whether jurors are able to generate coincidental match exemplars in some conditions but not others. If the incidence rate is lowered to the point where match exemplars cannot be generated, then we would expect that the presentation effects identified in Experiment 2 for target and frame will disappear.

With this in mind, Experiment 3 was designed to examine the effects of incidence rate variations on the exemplar pattern observed thus far.

Method

Participants

The participants were 440 jury-eligible undergraduate business students (52% female) at the University of Texas. Compensation and identification of jury eligibility were the same as in Experiment 2. Data from 23 non-U.S. citizens and one convicted felon were excluded from the original sample ($n = 464$).

Design and Predictions

Experiment 3 employed a fully crossed 2 (target: single, multi) × 2 (frame: probability, frequency) × 3 (incidence rate: 1/1,000, 1/1,000,000, 1/1,000,000,000) between-subjects design. Participants were assigned at random to one of the 12 conditions. The stimuli were similar to those used in Experiment 2. Jurors were asked to estimate P(source) and P(guilt) and to render a verdict.

Because the crime was said to occur in a large city (Houston), the natural reference class for imagining coincidental match exemplars is a population of one or more million people. For incidence rates of 1/1,000, coincidental match exemplars clearly exist and the theory predicts effects for target and frame similar to those observed in Experiment 2. For incidence rates of 1/1,000,000,000 (1 in 1 billion), coincidental match exemplars probably do not exist even for an urban reference class of several million people. Therefore, target and frame effects probably will not exist at this incidence rate. For incidence rates of 1/1,000,000, coincidental match exemplars may or may not exist depending on whether people believe that multiple people will match a 1 in 1,000,000 DNA profile in a large city.

Because target and frame effects may depend on incidence rates, target × incidence rate and frame × incidence rate interactions for all dependent measures were predicted. As the incidence rate gets smaller, exemplars may be hard to generate in any of the target and frame conditions and the effects for target and frame may be eliminated.

Results

As predicted, a MANOVA revealed a target × incidence rate, $F(4, 842) = 2.47$, $p = .043$, interaction and a frame × incidence rate interaction on P(source) and P(guilt), $F(4, 842) = 2.99$, $p = .018$. The effects of target and frame on P(source) and P(guilt) were significantly smaller (or eliminated altogether) for smaller incidence rates. Figure 1 presents the P(source) means for the target × incidence rate interaction (upper panel) and the frame × incidence rate interaction (lower panel). The target and frame effects observed in Experiment 2 were replicated on P(source) for the 1 in 1,000 incidence rate, target: single $M = 70.0\%$, multi $M = 50.9\%$, $F(1, 149) = 11.79$, $p < .001$; frame: prob. $M = 72.9\%$, freq. $M = 48.0\%$, $F(1, 149) = 19.91$, $p < .001$. However, these presentation effects were smaller for the 1 in 1,000,000 incidence rate, target: single $M = 80.6\%$, multi $M = 75.7\%$, $F(1, 131) = 1.05$, $p > .05$; frame: prob. $M = 82.2\%$, freq. $M = 74.1\%$, $F(1, 131) = 2.78$, $p = .098$, and nonexistent for the 1 in 1,000,000,000 incidence rate, target: single $M = 82.9\%$, multi $M = 83.1\%$, $F(1, 143) < 1$, $p > .05$; frame: prob. $M = 83.7\%$, freq. $M = 82.3\%$, $F(1, 143) < 1$, $p > .05$.

Patterns of extreme responses on the P(source) measure as a function of the four target/frame presentations (s/p, s/f, m/p, m/f) for each incidence rate appear in three panels in Fig. 2. This figure presents the proportion of jurors in each condition that gave extremely high (≥99%) and extremely low (≤1%) P(source) estimates.

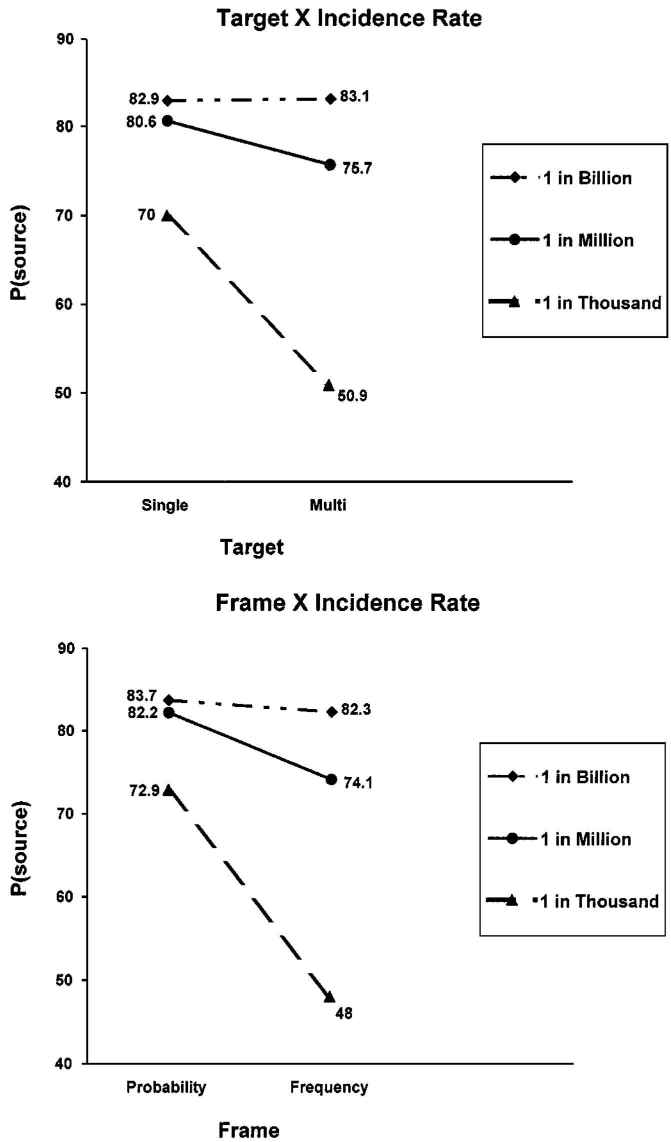

Fig. 1. P(source) means for target × incidence rate (top panel) and frame × incidence rate (bottom panel). Cell sizes in each panel range from 62 to 81 and 65 ot 78, respectively.

For jurors in the 1 in 1,000 incidence rate conditions, a noteworthy result appeared. A majority of jurors in the s/p condition (54.1%) reported that they were certain or nearly certain that the defendant was the source of the recovered blood evidence, P(source) $\geq$ 99%. However, more than one-third of jurors in the m/f condition (35.9%) were certain or nearly certain that the defendant was *not* the source of the

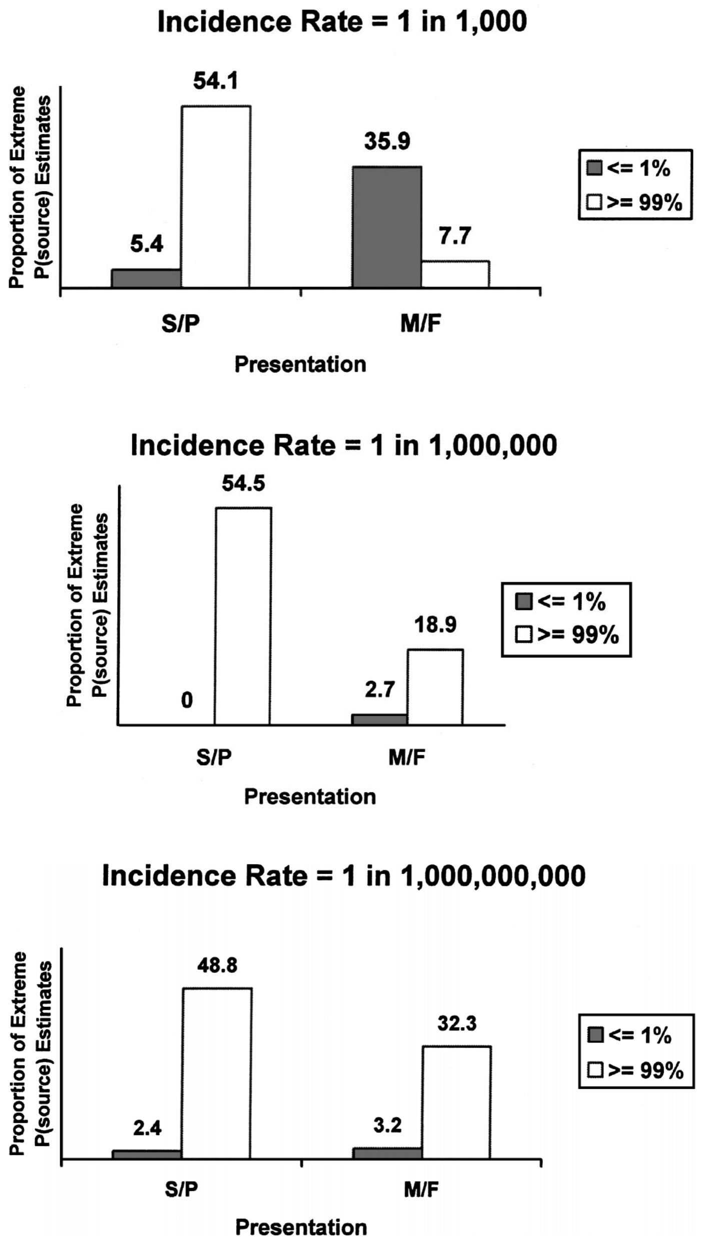

Fig. 2. Proportion of extreme P(source) estimates (≤1% and ≥99%) by jurors in the s/p (single-target, probability frame) and m/f (multi-target, frequency frame) conditions as a function of incidence rates. Cell sizes in each panel range from 37 to 39, 33 to 37, and 31 to 41, respectively.

blood evidence, P(source) $\leq$ 1%. This result is particularly striking when one bears in mind that the only difference in the stimuli received by jurors in these conditions is that the s/p group was told "The probability that the suspect would match the blood drops if he were not their source is 0.1%," whereas the m/f group was told "One in 1,000 people in Houston would also match the blood drops." The two lower panels of Fig. 2 show that this result did not emerge at smaller incidence rates. At incidence rates of 1 in 1,000,000 and 1 in 1,000,000,000, only about 3% of jurors gave P(source) responses $\leq$1%. However, even at these very low incidence rates, twice as many s/p jurors gave P(source) estimates $\geq$99% than m/f jurors (Ps = 51.4 and 25.0%, respectively; $z = 2.56$, $p < .006$).

Loglinear analyses revealed significant effects for target, Wald χ^2 (1, n = 434) = 9.73, p = .0018, and incidence rate, Wald χ^2 (1, n = 434) = 6.40, p = .0114, on Verdict. Jurors were about 50% more likely to convict when the DNA match statistic was presented as a single-target (36.5%) than when it was presented as a multi-target (23.2%). The main effect for incidence rate indicated that jurors convicted more frequently as the incidence rate became smaller (1/1,000 $P = 17.1\%$; 1/1,000,000 $P = 29.4\%$; 1/1,000,000,000 $P = 43.8\%$).

The data in Experiment 3 suggest that the target and frame effects observed earlier depend on incidence rate levels. Specifically, target and frame effects became smaller as incidence rates became smaller. This pattern is consistent with exemplar cueing theory because a juror's ability to generate coincidental match exemplars, even in exemplar-friendly conditions (e.g., the m/f condition), is harmed by extremely small incidence rates.

Experiment 3 also supported the proposition that there may be a ceiling effect on the perceived probative value of statistical DNA evidence. Even when DNA match incidence rates are extraordinarily small (e.g., 1 in 1 billion), there are limits on the amount of weight that jurors are willing to give this type of evidence. Although it is hard to say whether the ceiling effect presented itself at the right level, jurors' failure to distinguish between incidence rates of 1 in 1 million and 1 in 1 billion is normatively defensible. This is so because the chance that the DNA match was produced in error is several orders of magnitude larger than 1 in 1 million (Koehler et al., 1995).

GENERAL DISCUSSION AND CONCLUSION

This paper tested a theory for predicting when laypeople will be more and less impressed by statistical DNA evidence. The DNA context was chosen both because of its practical importance in today's courtroom environment and because it often involves very low probability matching evidence. In this setting, exemplar cueing theory suggests that there is an important psychological difference between (a) recognizing the possibility that a seemingly damning DNA match arose through an unlucky coincidence and (b) realizing that examples of such coincidental matches exist and are plentiful. In the first situation, a small, abstract chance may be treated as essentially zero (Stone, Yates, & Parker, 1994). In the second situation, exemplars transform mere statistical possibility into imagery that is more compelling

(Slovic et al., 2000). Exemplar cueing is probably less relevant in cases involving moderate and large probabilities because decision makers are less likely to treat such probabilities as essentially zero.

In the experiments reported here, the salience of coincidental match exemplars was influenced by varying the target (single person or broad class of people) and frame (probability or frequency) of the statistical evidence. The presentation effects were, at times, dramatic. For example, DNA evidence that convinced half of the s/p jurors that the suspect almost certainly was the source ($p \geq 99\%$), simultaneously convinced one-third of the m/f jurors that the suspect almost certainly was *not* the source ($p \leq 1\%$). Such results suggest that the persuasive power of DNA statistics may have as much to do with how those statistics are presented and framed as it does with the theoretical power of the technology.

The strategic implications of these data for trial attorneys and DNA experts seem clear. Other things being equal, the prosecution should present DNA statistics in a single-target, probability frame format (Koehler, in press). This presentation makes it difficult to take seriously the possibility that the match is merely coincidental. The defense should favor a multi-target, frequency frame format in cases where exemplar generation seems reasonable (i.e., where the incidence rate is not smaller than the reference class jurors will most likely use). Judges and legislators may also find this research useful when considering standards for presenting scientific and statistical evidence in court. For example, judicial instructions might be formulated that acknowledge that there are different ways of presenting the same statistical information. As the data from the dual presentation conditions data in Experiment 2 suggest, sensitizing legal decision makers to presentation effects may moderate those effects. The results of Experiment 3 suggest that strategic presentations of DNA statistics are less important when the incidence rates are so small relative to the reference class that exemplar generation is difficult under any target and frame formulation.

Exemplar generation may be influenced by variables other than target and frame. Koehler and Macchi (2001) showed that the defense-friendly m/f method for presenting DNA statistics could be turned into a prosecution-friendly method by blocking people's cognitive access to exemplars. This was accomplished by offering mock jurors one of the three mathematically equivalent frequency statistics for each of two different incidence rates. For a 1 in 1,000 incidence rate, the statistics were presented as "1 out of 1,000," "0.1 out of 100," or "2 out of 2,000." For a 1 in 100,000 incidence rate, the statistics were presented as "1 out of 100,000," "0.1 out of 10,000," or "2 out of 200,000." Koehler and Macchi (2001) found that people were substantially more persuaded by the statistical evidence when it appeared with a fractional numerator. They argued that the near-zero numerators implied that no one else is likely to match (i.e., zero exemplars) and otherwise made it harder for people to think about others who might match.

Future researchers may wish to explore other aspects of exemplar cueing. Are there other ways to cue exemplars? When are the effects likely to persist? Can the effects be altered by such variables as repetition (Arkes, 1993), cognitive overload (Gilbert & Osborne, 1989), statistical training (Nisbett, Fong, Lehman, & Cheng, 1987), or group deliberation (McCoy, Nunez, & Dammeyer, 1999)? Future researchers may also wish to introduce more complexity and variety into the experimental stimuli to determine whether the exemplar effects identified here affect various real-world decision making.

APPENDIX: SAMPLE STIMULUS: EXPERIMENT 2 (MULTI-TARGET, FREQUENCY FRAME)

In the case of *State of Texas v. Nethers*, Steven Nethers was accused of murdering Richard Oden during an attempted robbery of a hardware store owned by Mr. Oden in Houston, Texas.

According to reliable eyewitness accounts, the perpetrator entered Mr. Oden's hardware store at approximately 4:30 P.M. on November 2, 1997, wearing a Halloween-type of mask and waving a small caliber handgun. The perpetrator approached Mr. Oden (who was behind the cash register) and said "Open it fast or you're a dead man."

According to the eyewitnesses, when the perpetrator turned his head to survey the store, Mr. Oden grabbed a hammer from the counter and smashed the perpetrator on head with a single blow. The perpetrator fired a single shot into Mr. Oden's chest and fled the store. Mr. Oden died shortly thereafter in a Houston hospital.

During an investigation of the hardware store crime scene, Houston police identified and recovered several moist blood drops from the path that was taken by the perpetrator as he fled the store. These drops were subjected to a form of DNA analysis called PCR testing. The PCR tests revealed the blood to be of a type known as "2, 3." Because this blood type was different from Mr. Oden's blood type, police believed that the recovered blood drops came from the bleeding head of the robber.

During routine interviews of people who live in neighborhood, the police identified several potential suspects. All of these individuals agreed to provide blood samples to police for comparison with blood that was recovered from the crime scene. One of the suspects, Mr. Steven Nethers, matched the 2, 3 blood type and was arrested for the murder.

At trial, the prosecution alleged that the blood analysis demonstrated Mr. Nethers was the source of the wet blood drops and that he was therefore guilty of attempted robbery and murder.

A DNA expert testified that his tests could not rule out Mr. Nethers as a possible source of the blood drops. He also testified that 1 in 1,000 people in Houston who are not the source would nonetheless match the blood drops.

The defense argued that the blood evidence is irrelevant because there was no direct evidence (such as eyewitness identifications) that linked Mr. Nethers to these crimes.

Questions

1. What would you say is the *numerical probability* that Mr. Nethers (the defendant) is the source of the blood evidence? _______%
2. What would you say is the *numerical probability* that Mr. Nethers (the defendant) murdered Mr. Oden (the victim)? _______%

3. As a juror, you are instructed to consider all of the evidence and arguments in this case carefully. You are to find against Mr. Nethers only if the evidence convinces you "beyond a reasonable doubt" that Mr. Nethers is guilty of this crime. What verdict would you return? (Please circle)

Not Guilty or Guilty?

4. Suppose this crime occurred in Austin, Texas (population: approximately 500,000). If every person in Austin were tested, approximately how many people who are not the source of the blood evidence would give a DNA match with the recovered blood evidence? (Please provide a *numerical* estimate): ______

ACKNOWLEDGMENTS

I thank Hal Arkes, Daniel Gilbert, Laura Macchi, Molly Mercer, Paul Slovic, Pat West, Paul Windschitl, members of the Behavioral Economics seminar at Harvard University, and students in the STA380 seminar at The University of Texas for comments about this research. I also thank Arienne Brint, Andrew Gershoff, Nir Liberboim, Caryn Thesing, and Cherilyn Young for superb research assistance.

This research was supported in part by a grant from the National Science Foundation, Decision, Risk and Management Science Program (SBR 9819305).

REFERENCES

ABC News (1998, January 30). World News Tonight with Peter Jennings.

Aitken, C. G. G. (1995). *Statistics and the evaluation of evidence for forensic scientists.* Chichester: Wiley.

Arkes, H. R. (1993). Some practical judgment and decision-making research. In N. J. Castellan, Jr. (Ed.), *Individual and group decision making: Current issues* (pp. 3–18). Hillsdale, NJ: Erlbaum.

Arkes, H. R., Connolly, T., & Hammond, K. R. (Eds.). (2000). *Judgment and decision making: An interdisciplinary reader.* Cambridge: Cambridge University Press.

CBS News (1998, January 29). Evening news with Dan Rather.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58,* 1–73.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, *66,* 819–829.

Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *62,* 328–339.

Evett, I. W., & Weir, B. S. (1998). *Interpreting DNA evidence: Statistical genetics for forensic scientists.* Sunderland, MA: Sinauer Associates.

Faigman, D. L., & Baglioni, A. J., Jr. (1988). Bayes' Theorem in the trial process. *Law and Human Behavior*, *12,* 1–17.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13,* 1–17.

Fienberg, S. E. (Ed.). (1989). *The evolving role of statistical assessments as evidence in the courts.* New York: Springer-Verlag.

Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. Cummins & C. Allen (Eds.), *The evolution of mind* (pp. 9–29). New York: Oxford University Press.

Gigerenzer G., & Hoffrage U. (1995). How to improve Bayesian reasoning without instructions: Frequency formats. *Psychological Review*, *102,* 684–704.

Gilbert, D. T., & Osborne, R. E. (1989). Thinking backward: Some curable and incurable consequences of cognitive busyness. *Journal of Personality and Social Psychology*, *57,* 940–949.
Goodman, J. (1986). *Probabilistic scientific evidence: Jurors' inferences.* Doctoral Dissertation, University of Washington, Ann Arbor, MI (University Microfilms no. 8706571).
Goodman, J. (1992). Jurors' comprehension and assessment of probabilistic evidence. *American Journal of Trial Advocacy*, *16,* 361–389.
Hansen, D. E., & Helgeson, J. G. (1996). The effects of statistical training on choice heuristics in choice under uncertainty. *Journal of Behavioral Decision Making*, *9,* 41–57.
Harries, C., & Harvey, N. (2000). Are absolute frequencies, relative frequencies, or both effective in reducing cognitive biases? *Journal of Behavioral Decision Making*, *13,* 431–444.
Hogarth, R. M. (Ed.). (1991). *Judgment and choice* (2nd ed.). Chichester: Wiley.
Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24,* 1–55.
Kahneman, D., & Lovallo, D. (1993). Bold forecasting and timid decisions: A cognitive perspective on risk taking. *Management Science*, *39,* 17–31.
Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristic and biases*. Cambridge: Cambridge University Press.
Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11,* 143–157.
Kaye, D. H. (1993). DNA evidence: Probability, population genetics, and the courts. *Harvard Journal of Law and Technology*, *7,* 101–172.
Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society A*, *154,* 75–81.
Kaye, D. H., & Sensabaugh, G. F. (2000). Reference guide on DNA evidence. In *Reference manual on scientific evidence* (2nd ed., pp. 485–576). Washington, DC, Federal Judicial Center.
Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *63,* 534–544.
Klar, Y., Medding, A., & Sarel, D. (1996). Nonunique invulnerability: Singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. *Organizational Behavior and Human Decision Processes*, *67,* 229–245.
Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios and error rates. *University of Colorado Law Review*, *67,* 859–886.
Koehler, J. J. (1997). One in millions, billions and trillions: Lessons from People v. Collins (1968) for People v. Simpson (1995). *Journal of Legal Education*, *47,* 214–223.
Koehler, J. J. (in press). The psychology of numbers in the courtroom: How to make DNA match statistics seem impressive or insufficient. *Southern California Law Review*.
Koehler, J. J., Chia, A., & Lindsey, J. S. (1995). The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial? *Jurimetrics Journal*, *35,* 201–219.
Koehler, J. J., & Macchi, L. (2001). *Exemplar cueing theory.* Unpublished manuscript.
Lander, E. S., & Budowle, B. (1994). DNA fingerprinting dispute laid to rest. *Nature*, *371,* 735–8.
Lempert, R. O. (1991a). Some caveats concerning DNA as criminal identification evidence: With thanks to the Reverend Bayes. *Cardozo Law Review*, *13,* 303–341.
Lempert, R. O. (1991b). Why do jury research? In R. Hastie (Ed.), *Inside the juror* (pp. 242–254). Cambridge: Cambridge University Press.
McCoy, M. L., Nunez, N., & Dammeyer, M. M. (1999). The effect of jury deliberations on jurors' reasoning skills. *Law and Human Behavior*, *23,* 557–575.
Miller, D. T., Turnbull, W., & McFarland, C. (1989). When coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology*, *57,* 581–589.
Mueller, L. D. (1993). The use of DNA typing in forensic science. *Accountability in Research*, *3,* 55.
Nance, D. (2000). *An empirical examination of the evidential use of likelihood ratios.* Unpublished manuscript.
National Research Council. (1992). *DNA technology in forensic science*. Washington, DC: National Academy Press.
National Research Council. (1996). *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press.
New York v. Wesley, 533 N.Y.S.2d 643, 644 (S. Ct. 1988).
Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, *238,* 625–631.
Reeves, T., & Lockhart, R. S. (1993). Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology: General*, *122,* 207–226.

Robertson, B., & Vignaux, G. A. (1995). *Interpreting evidence: Evaluating forensic science in the courtroom.* Chichester: Wiley.

Rowe, B. I. (1997). A possible solution for the problem of juries slighting nonscientific evidence: A Bayesian-like judicial instruction. *American Journal of Criminal Law*, *24,* 541–579.

Saks, M. J., & Kidd, R. F. (1980/1981.). Human information processing and adjudication: Trial by heuristics. *Law and Society Review*, *15,* 123–160.

Schklar, J., & Diamond, S. (1999). Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior*, *23,* 159–184.

Shepard, S. (1998, January 31). That mystery dress: An example of how rumors pass for facts. *Houston Chronicle*, p. A6.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119,* 3–22.

Slovic, P., & Monahan, J. (1995). Probability, danger and coercion: A study of risk perception and decision making in mental health law. *Law and Human Behavior*, *19,* 49–65.

Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, *24,* 271–296.

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.

Stone, E. R., Yates, F. J., & Parker, A. M. (1994). Risk communication: Absolute versus relative expressions of low-probability risks. *Organizational Behavioral and Human Decision Processes*, *60,* 387–408.

Taroni, F., & Aitken, C. G. G. (1998a). Probabilistic reasoning in the law: Part 2. Assessment of probabilities and explanation of the value of trace evidence other than DNA. *Science & Justice*, *38,* 179–188.

Taroni, F., & Aitken, C. G. G. (1998b). Probabilistic reasoning in the law: Part I. Assessment of probabilistic and explanation of the value of DNA evidence. *Science and Justice*, *38,* 165–177.

Thompson, W. C. (1989). Are juries competent to evaluate statistical evidence? *Law and Contemporary Problems*, *52,* 9–41.

Thompson, W. C., Britton, R. L., & Schumann, E. L. (1990). *Jurors' sensitivity to variations in statistical evidence.* Unpublished manuscript.

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutors' fallacy and the defense attorney's fallacy. *Law and Human Behavior*, *11,* 167–187.

Teigen, K. H., Brun, W., & Frydenlund, R. (1999). Probability and risk judgments: The role of frequentistic information. *Journal of Behavioral Decision Making*, *12,* 123–139.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5,* 207–232.

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

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

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