Research Article

Thinking About Low-Probability Events

An Exemplar-Cuing Theory

Jonathan J. Koehler[1] **and Laura Macchi**[2]

[1]*McCombs School of Business and School of Law, The University of Texas at Austin, and* [2]*Università degli Studi di Milano-Bicocca, Milan, Italy*

ABSTRACT—***The way people respond to the chance that an unlikely event will occur depends on how the event is described. We propose that people attach more weight to unlikely events when they can easily generate or imagine examples in which the event has occurred or will occur than when they cannot. We tested this idea in two experiments with mock jurors using written murder scenarios. The results suggested that jurors attach more weight to the defendant's claim that an incriminating DNA match is merely coincidental when it is easy for them to imagine other individuals whose DNA would also match than when it is not easy for them to imagine such individuals. We manipulated the difficulty of imagining such examples by varying the description of the DNA-match statistic. Some of the variations that influenced the jurors were normatively irrelevant.***

The way people respond to probabilistic information depends on how that information is communicated. For example, Slovic, Monahan, and MacGregor (2000) demonstrated that clinicians were more concerned about the recidivism risk posed by a mental patient when the risk was presented as a frequency (e.g., 1 in 20) than when it was presented as a probability (.05). The authors argued that this occurred because the frequency presentation created "frightening images" of recidivists in the minds of the clinicians. Although this explanation may fit the data Slovic et al. obtained, the thoughts or images that statistical representations promote need not be frightening or otherwise arousing to influence decision makers. It may be enough that some representations evoke thoughts of other examples in which the event has occurred or will occur.

For example, suppose that a woman is considering purchasing a lottery ticket. We suggest that inducing her to think about other lottery tickets that were winners will make her more inclined to make the purchase, even if these thoughts do not alter her subjective probabilities of winning. The effect of such thoughts is probably greatest for very low-probability events because in such cases the thoughts may cause an event that would otherwise be ignored to receive attention. With this idea in mind, we offer an exemplar-cuing theory for how people assign weight to low-probability events. This theory proposes that the weight decision makers attach to low-probability events is, in part, a function of whether they can easily generate or imagine exemplars (examples) for the event. Furthermore, the manner in which the events are described affects whether or not exemplars are imagined (Koehler, 2001).

As an illustration, consider two descriptions of the chance of winning the daily three-digit New York Lottery Numbers game: (a) "There is a 0.1% chance that a given ticket will win," and (b) "One in every 1,000 tickets out of the 500,000 tickets that are sold each day will win." Both descriptions are accurate, but only the latter cues the potential ticket purchaser to think about more than one ticket that might win each day.

The availability of examples in the decision maker's mind is important to exemplar-cuing theory. However, the term "availability" is used differently in this theory than in research on the availability heuristic. The availability heuristic holds that people assess the probability of an event on the basis of the ease with which they can imagine relevant instances (Tversky & Kahneman, 1973). In contrast, exemplar availability is an indicator of the weight that people attach to low-probability events whose values are not in question.

THEORETICAL DEVELOPMENT

Exemplar-cuing theory is informed by research on *ratio bias* and the probability-frequency distinction. Ratio-bias studies find that when the objective probability of winning is identical across lotteries, most people prefer playing a large lottery (e.g., 10 winning jelly beans out of 100) over playing a small lottery (e.g., 1 winning jelly bean out of 10; Kirkpatrick & Epstein, 1992; Miller, Turnbull, & McFarland, 1989; Windschitl & Wells, 1996). Indeed, people prefer to play a large lottery even when a small lottery offers slightly more favorable odds (Denes-Raj & Epstein, 1994; Pacini & Epstein, 1999).

Epstein and his colleagues use two principles from cognitive-experiential self-theory (see Kirkpatrick & Epstein, 1992) to explain

Address correspondence to Jonathan J. Koehler, CBA 5.202, The University of Texas at Austin, Austin, TX 78712-1175; e-mail: koehler@mail.utexas.edu.

Journal: PSCI ☑ Disk used ED: Saravanan Pgn by: Anand kumar
Article: 716 Pages:7 (col.fig.: NIL) Despatch Date: 25/5/2004

ratio bias. The first principle states that large ratios (e.g., 10/100) are more difficult to comprehend than small ratios (e.g., 1/10). The second principle states that people attend more closely to the absolute number of ways an outcome can occur than to probability ratios (Pacini & Epstein, 1999). Miller et al. (1989) also pointed to the importance of the absolute number of ways an outcome can occur, but they drew on norm theory (Kahneman & Miller, 1986) for an explanation of ratio bias. They suggested that winning a large lottery seems less suspicious, hence more "normal," than winning a small lottery that has identical odds. In five experiments, they found that "the fewer ways a low probability event could have occurred by chance, the less inclined subjects were to assume that the event occurred by chance" (p. 584). Exemplar-cuing theory can also account for ratio bias, though it does not require a comparison between options that differ in the absolute number of ways a favorable outcome can occur. Instead, exemplar cuing requires only consideration of whether examples of other "winners" are cognitively available, regardless of the actual number of ways that winners might appear by chance.

Research on the probability-frequency distinction also informs our theory. Gigerenzer and Hoffrage (1995) compared "frequency formats" with "single event probabilities" and concluded that people make more accurate judgments with frequency formats (see also Brase, Cosmides, & Tooby, 1998; Cosmides & Tooby, 1996). According to Gigerenzer and Hoffrage, single-event probabilities appear "in the form of probabilities, attached to a single person" (p. 686). In terms of our lottery example, "there is a 0.1% probability that Elizabeth's lottery ticket will win" is a single-event probability. In contrast, "one in 1,000 tickets sold in the lottery will win" illustrates a frequency format. Notice that the two statements differ in both statistical *form* (0.1% vs. 1 in 1,000) and *target* (Elizabeth's single lottery ticket vs. multiple tickets sold in the lottery). Because single-event probabilities and frequency formats differ in two respects, it is not obvious which factor—form or target—is responsible for previously observed effects when these formats were compared. We consider the role of both factors next.

Statistical Form and Target

Both the target (single vs. multiple) and the form (probability vs. frequency) of a probabilistic statement may affect the production of exemplars and the weights people assign to low probabilities. However, we suspect that target—despite its near invisibility in the literature (cf. Jones, Jones, & Frisch, 1995; Macchi, 1995)—may be the more important variable.

The target of a probabilistic statement identifies a problem-relevant sample space or reference class. Single targets (e.g., "Elizabeth's lottery ticket") offer the smallest reference class ($n = 1$) by focusing attention narrowly on one event. This focus discourages exemplar production because there are no other events to consider. Consequently, in the lottery example, people who receive single targets are unlikely to think about other lottery tickets that could win. In contrast, multiple targets (e.g., "tickets sold in the lottery") promote exemplar production because they offer a larger reference class ($n > 1$) within which to consider other winning tickets.

However, the mere presence of a multiple target is not sufficient to guarantee the production of exemplars. Whether exemplars are cued or not is determined by the combination of (a) the size of the reference class (as identified by the target) and (b) the rareness of the event. In the lottery example, if a multiple target is provided (e.g., 500,000 tickets are sold) and the chance that a particular ticket will win is 0.1%, then winning-ticket exemplars are likely to be cued (because approximately 500 tickets in this lottery will be winners). However, if the multiple target indicates that only 5 tickets are sold, then exemplars are not likely to be cued. Algebraically, if the product of the reference-class size and the event probability is greater than 1, exemplars are likely to be cued. Cuing of exemplars increases the weight people attach to the possibility that the unlikely coincidental outcome will occur (e.g., Elizabeth's lottery ticket wins). However, when the product is less than 1, exemplars are unlikely to be cued, and people will give little weight to the possibility of winning.

Unlike the target, the form (probability or frequency) of a probabilistic statement does not identify a relevant sample space within which to search for exemplars. Therefore, we predict that variations in form will not influence decision makers' probability judgments.[1] This prediction is consistent with results obtained by Macchi (2000), which showed no effect of form on judgment when the effects of target and other variables were controlled.

Exemplars in the Courtroom

We tested exemplar-cuing theory in a legal decision-making context, in part because the way people think about probabilistic evidence is a significant concern in the courtroom. Scientists increasingly testify about low probabilities in cases involving discrimination, deceptive trade practices, and forensic science (Gastwirth, 2000). However, little is known about the impact of this testimony on jurors (National Research Council, 1996, pp. 203–204).

Perhaps the most important and pervasive probabilistic evidence arises in cases involving DNA matches between a criminal suspect and trace evidence from a crime scene. Consider the following scenario:

> A masked man bursts into a hardware store and announces his intent to rob it. However, the owner of the store resists and punches the robber in the nose. The robber flees the store, but not before a small drop of blood falls from his nose. DNA analysis of the blood reveals a match with a saliva sample taken from one of several suspects. The matching suspect is charged with attempted robbery and brought to trial.

In such a case, a scientist will testify about the DNA match and identify the frequency of this DNA profile among humans. This value, which is often 1 in thousands, millions, or billions (Koehler, 1997), is the *incidence rate*. Defendants often attempt to rebut DNA evidence by claiming that the match is merely coincidental. Exemplar-cuing theory predicts that the weight jurors assign to this claim depends not merely on the value of the incidence rate, but also on whether the DNA evidence is presented in a way that cues exemplars (i.e., examples of other people who might match the DNA). The theory also predicts that, for a fixed incidence rate, exemplars are more likely to be cued when the product of the incidence rate and the reference-class size exceeds 1 than when this product is less than 1. Thus, people will attach more weight to the defendant's claim that the DNA

[1]The frequency form may cue exemplars if the sample space implied by the frequency induces people to adopt a broader search strategy.

TABLE 1
The Eight Conditions in Experiment 1

Target	Form	Reference class size	Description of DNA-match statistic	Exemplar state & prediction
Single	Probability	Small (500)	The chance that the suspect would match the blood drops if he were not their source is 0.001%.	No exemplars; evidence is more convincing
Single	Probability	Large (5,000,000)	The chance that the suspect would match the blood drops if he were not their source is 0.001%.	No exemplars; evidence is more convincing
Single	Frequency	Small (500)	The chance that the suspect would match the blood drops if he were not their source is 1 in 100,000.	No exemplars; evidence is more convincing
Single	Frequency	Large (5,000,000)	The chance that the suspect would match the blood drops if he were not their source is 1 in 100,000.	No exemplars; evidence is more convincing
Multiple	Probability	Small (500)	0.001% of the people in a town who are not the source would nonetheless match the blood drops.	No exemplars; evidence is more convincing
Multiple	Probability	Large (5,000,000)	0.001% of the people in a town who are not the source would nonetheless match the blood drops.	Exemplars; evidence is less convincing
Multiple	Frequency	Small (500)	1 in 100,000 people in a town who are not the source would nonetheless match the blood drops.	No exemplars; evidence is more convincing
Multiple	Frequency	Large (5,000,000)	1 in 100,000 people in a town who are not the source would nonetheless match the blood drops.	Exemplars; evidence is less convincing

match is coincidental when this product is greater than 1 than when it is not. This weighting would be reflected in greater skepticism about the strength of the prosecution's case against the defendant. We investigated these ideas in two legal decision-making experiments.

EXPERIMENT 1

Method

Participants

Four hundred forty-one undergraduate and graduate students at an Italian University participated in exchange for credit in introductory psychology classes. Participant-jurors completed the study in Italian.

Design and Procedure

The study employed a fully crossed 2 (target: single, multiple) × 2 (form: probability, frequency) × 2 (reference-class size: small, large) between-subjects design. Jurors were provided with a 2-page summary of a murder case in which DNA evidence recovered from the crime scene matched the defendant (incidence rate = 1 in 100,000). To avoid ceiling effects on the dependent measures, we provided little other evidence against the defendant. Case materials were identical across conditions except for (a) the wording of the DNA match-statistic and (b) the size of the reference class. We manipulated reference-class size by identifying the size of the city in which investigators believed the murderer resided (small population = 500; large population = 5,000,000). The incidence rate and population sizes were selected such that exemplars of a low-incidence-rate event would be available for a large reference class only. Details about the eight conditions appear in Table 1.

Jurors estimated the strength of the evidence against the defendant (evidence strength); the probability that the defendant was the source of the DNA evidence, *P*(source); and the probability that the defendant committed the crime, *P*(guilt). They made their judgments of evidence strength on a 7-point Likert-type scale (1 = *not at all strong*, 7 = *extremely strong*), and generated their own values for the probability questions. Jurors also provided verdicts[2] and written explanations.

Our theory predicted that exemplars were most likely to be cued among jurors who received the DNA statement with both a multiple target and a large reference class. We expected jurors in these conditions to attach relatively more weight than other jurors to the defendant's claim that the DNA match was coincidental. This outcome would be reflected in relatively fewer guilty verdicts and lower values for evidence strength, *P*(source), and *P*(guilt). In statistical terms, we predicted a Target × Reference Class interaction for all dependent measures.

Results

The results for all eight conditions appear in Table 2. As predicted, we found a Target × Reference Class interaction for evidence strength, $F(1, 433) = 5.51$, $p = .019$; *P*(source), $F(1, 402) = 13.26$, $p < .001$;

[2] Jurors were instructed to find the defendant guilty "only if the evidence convinces you 'beyond a reasonable doubt' that [the defendant] is guilty of this crime" (English translation).

TABLE 2

Mean Judgments as a Function of Condition in Experiment 1

Condition (target, form, reference-class size)	Judgment: Evidence strength (1–7)	P(source)	P(guilt)	Verdict of guilty (%)
Single target, probability, small class	4.0 (0.2)	.58 (.05)	.51 (.05)	28 (6)
Single target, probability, large class	4.3 (0.2)	.63 (.05)	.62 (.05)	21 (5)
Single target, frequency, small class	4.0 (0.2)	.47 (.05)	.52 (.05)	24 (6)
Single target, frequency, large class	4.1 (0.2)	.55 (.05)	.54 (.05)	20 (6)
Multiple targets, probability, small class	4.0 (0.2)	.55 (.05)	.52 (.05)	23 (6)
Multiple targets, probability, large class	3.7 (0.2)	.37 (.05)	.38 (.04)	10 (4)
Multiple targets, frequency, small class	3.8 (0.2)	.50 (.05)	.47 (.05)	20 (6)
Multiple targets, frequency, large class	3.4 (0.2)	.31 (.05)	.32 (.05)	9 (4)

Note. Standard errors are in parentheses. Cell sizes ranged from 101 to 118. *P*(source) = probability that the defendant was the source of the DNA evidence; *P*(guilt) = probability that the defendant committed the crime.

P(guilt), $F(1, 402) = 10.67$, $p = .001$; and verdict, $\chi^2(1, N = 428) = 11.08$, $p = .002$. As Table 3 shows, all four of our dependent measures indicate that jurors in the multiple-target, large-reference-class conditions—the only exemplar-conducive conditions—were less impressed by the evidence than were jurors in the other conditions.

The written explanations, which were coded by trained, hypothesis-blind Italian judges, revealed a higher proportion of exemplar-related comments among jurors in the exemplar-conducive conditions (i.e., multiple target and large reference class) than in the other conditions ($Z = 3.12$, $p < .001$). This result is consistent with the Target × Reference Class interaction.

As expected, we did not detect a significant Form × Reference Class interaction for evidence strength, *P*(source), *P*(guilt), or verdict. However, we detected a marginally significant main effect for form, $F(2, 401) = 2.58$, $p = .077$, on the probability judgments. Jurors in the frequency-form conditions assigned lower *P*(source) and *P*(guilt) values than did jurors in the probability-form conditions.

Discussion

The results supported our predictions and provide insight into when and why people attach weight to low-probability events (such as a coincidental DNA match). When low-probability evidence was presented in an exemplar-conducive way, jurors were concerned about the possibility that the match was coincidental and generally did not believe that the defendant was the source of the DNA (median = 25%) or guilty of the crime (median = 32%). When evidence was not presented in an exemplar-conducive manner, jurors were apparently less concerned about the possibility of a coincidence, and therefore more likely to believe that the defendant was the source of the DNA (median = 55%) and guilty of the crime (median = 55%).

We suggest that this difference emerged because a 1-in-100,000 incidence rate in combination with a 5,000,000-member reference class helped trigger thoughts about people other than the defendant who would also match the DNA evidence. This theory received additional support from the jurors' written explanations, which showed that jurors in the exemplar-conducive conditions expressed the most concern about others who might match the evidence.

Statistical form (probability vs. frequency) did not interact with reference-class size to affect judgments, and there was a marginally significant main effect for form on two of the four dependent measures. The lack of a robust effect for form may seem surprising in light of frequency effects observed in other studies (Cosmides & Tooby, 1996; Slovic et al., 2000; Thompson & Schumann, 1987). However, most of these effects were obtained with incidence rates substantially larger than those used here (but see Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000).

Experiment 2

When low-probability evidence appears with a large reference class, exemplar-cuing predictions are consistent with Bayesian predictions. This fact makes it difficult to know, in such cases, whether people are reasoning about low-probability events by exemplars or by the nor-

TABLE 3

Mean Judgments as a Function of Target and Reference-Class Size (Collapsed Across Form) in Experiment 1

Target and reference-class size	Judgment: Evidence strength (1–7)	P(source)	P(guilt)	Verdict of guilty (%)
Single target, small class	4.0 (0.1)	.53 (.03)	.51 (.03)	26 (4)
Single target, large class	4.2 (0.1)	.59 (.04)	.58 (.03)	20 (4)
Multiple targets, small class	3.9 (0.1)	.52 (.03)	.50 (.03)	22 (4)
Multiple targets, large class	3.6 (0.1)	.34 (.03)	.35 (.03)	9 (3)

Note. Standard errors are in parentheses. Cell sizes ranged from 101 to 118. *P*(source) = probability that the defendant was the source of the DNA evidence; *P*(guilt) = probability that the defendant committed the crime.

TABLE 4
Mean Judgments as a Function of Condition in Experiment 2

Condition	Judgment			
	Evidence strength (1–7)	*P*(source)	*P*(guilt)	Verdict of guilty (%)
Higher incidence rate (.001)				
0.1 exemplar	4.5 (0.3)	.63 (.07)	.56 (.06)	24 (8)
1 exemplar	3.5 (0.2)	.43 (.07)	.39 (.07)	16 (7)
2 exemplars	3.6 (0.2)	.24 (.07)	.26 (.06)	13 (6)
Lower incidence rate (.00001)				
0.1 exemplars	4.5 (0.3)	.72 (.07)	.67 (.07)	35 (9)
1 exemplar	4.4 (0.2)	.53 (.07)	.52 (.07)	17 (7)
2 exemplars	3.9 (0.3)	.46 (.08)	.43 (.06)	7 (5)

Note. Standard errors are in parentheses. Cell sizes ranged from 28 to 31. *P*(source) = probability that the defendant was the source of the DNA evidence; *P*(guilt) = probability that the defendant committed the crime.

mative rules of probability. Consider a criminal case in which a DNA match is the only evidence against a suspect and the product of the reference-class size and the incidence rate exceeds 1 (e.g., $1{,}000{,}000 \times 1/100{,}000 = 10$). Exemplar cuing predicts that people will be relatively unpersuaded by the statistical evidence because exemplars are readily available. Similarly, a Bayesian analysis indicates that the suspect probably is not the source of the DNA evidence.[3]

In Experiment 2, we pitted the predictions of exemplar-cuing theory against those of Bayesian theory in a legal decision-making task by varying the numerators and denominators—but not the ratios—of the incidence rate. Exemplar-cuing theory predicts that in our courtroom scenario, incidence rates that contain fractional numerators (e.g., "0.1 out of 100") inhibit exemplars and thereby appear to weaken the defendant's claim that the DNA match is coincidental. Incidence rates that contain numerators greater than or equal to 1 (e.g., "1 out of 1,000" or "2 out of 2,000") provide readily available exemplars (from the numerator of the incidence rate itself) and thereby appear to strengthen the defendant's claim. The Bayesian predictions do not vary across the exemplar conditions just described. That is, according to Bayesian theory, the weight that people assign to the defendant's claim of a coincidental match should not depend on which numerators and denominators are used to describe mathematically identical incidence rates.

Method

Participants
One hundred seventy-eight University of Texas students participated in exchange for course credit.

Design and Procedure
Case materials were similar to those in the multiple-target, frequency-form conditions from Experiment 1, except that they were in English and did not include information about the size of the town in which the perpetrator resided. We also introduced two incidence rates to test the robustness of the exemplar-cuing predictions.

The study employed a fully crossed 2 (incidence rate: .001, .00001) × 3 (exemplars in numerator: 0.1, 1, 2) between-subjects design. When the incidence rate was .001, it appeared as "0.1 out of 100," "1 out of 1,000," or "2 out of 2,000." When the incidence rate was .00001, it appeared as "0.1 out of 10,000," "1 out of 100,000," or "2 out of 200,000." Exemplar-cuing theory predicts that incidence rates that include at least one exemplar (e.g., 1 out of 1,000 or 2 out of 2,000) will call jurors' attention to other individuals who might also match. This effect will produce concern about a coincidental match, and jurors will be relatively less impressed by the evidence. In contrast, incidence rates that include a fractional number of exemplars (e.g., 0.1 out of 100) will discourage thoughts about exemplars, and jurors will be more convinced by the prosecution's evidence. As a secondary matter, we predicted that jurors would be more convinced by the prosecution's case when the incidence rate was smaller (i.e., more diagnostic) than when it was larger.

Results

The results supported the prediction of exemplar-cuing theory (see Table 4). We found a main effect for number of exemplars on evidence strength, $F(2, 172) = 4.36$, $p = .014$; P(source), $F(2, 172) = 11.18$, $p < .001$; P(guilt), $F(2, 172) = 8.85$, $p < .001$; and verdict, $\chi^2(2, N = 178) = 7.29$, $p = .026$. As the number of exemplars increased from 0.1 to 1 to 2, the jurors judged the statistical evidence and case to be progressively weaker and were less likely to convict. These effects were driven largely by the fractional conditions.[4] Post hoc contrasts using Tukey's HSD test revealed higher P(source) and P(guilt) estimates in the 0.1-exemplar conditions than in the 1-exemplar conditions, $p = .025$ for P(source) and $p = .062$ for P(guilt), or

[3]According to the odds form of Bayes' theorem, $P(\text{source} \mid \text{match})$ may be calculated as follows:

$$\frac{P(\text{source}|\text{match})}{P(\text{not source}|\text{match})} = \frac{P(\text{source})}{P(\text{not source})} \times \frac{P(\text{match}|\text{source})}{\text{P}(\text{match}|\text{not source})} = \frac{.000001}{.99999} \times \frac{1}{.00001}.$$

Converting these odds back to probabilities yields $P(\text{source} \mid \text{match}) = .091$.

[4]A reviewer wondered whether our results in the fractional conditions sprang from confusion about the underlying ratios. However, there is evidence that the jurors understood the fractional ratios. First, jurors who received the more probative fractional ratio (i.e., 0.1/10,000) properly judged the case to be stronger than those who received the less probative fractional ratio (i.e., 0.1/100). Second, in another study (Koehler, 2002), we observed that jurors who heard a .00001 incidence rate described as either 0.001/100 or 0.1/10,000 treated these ratios as equally compelling proof. This result is predicted by both the normative theory and exemplar-cuing theory.

the 2-exemplar conditions, $p < .001$ for P(source) and $p < .001$ for P(guilt). Post hoc contrast analyses did not reveal significant differences between the 1-exemplar and 2-exemplars conditions on any dependent measure.[5]

The results partially supported our secondary prediction regarding incidence rate. We detected a main effect for incidence rate on P(source), $F(1, 167) = 4.96$, $p = .027$, and P(guilt), $F(1, 167) = 6.93$, $p = .009$. Jurors assigned higher estimates when the incidence rate was more diagnostic (i.e., .00001) than when it was less diagnostic (i.e., .001). The results for evidence strength and verdict were in the predicted direction but did not reach statistical significance.

Discussion

The results of Experiment 2 provide further support for the influence of exemplar cuing on people's evaluations of low-probability events. When evidence was provided with a multiple target and in frequency form, jurors treated the evidence as more compelling when the numerator was a fraction than when it was larger. Presumably, fractional values discouraged jurors from thinking about other people who would match the DNA by coincidence, whereas larger numerators provided readily available exemplars. Indeed, the exemplar effect was strong enough that jurors were sometimes more convinced by relatively weak evidence that did not promote exemplars than by stronger evidence that did promote exemplars. As Table 4 shows, jurors treated a .001 incidence rate described as 0.1/100 as stronger evidence than a .00001 incidence rate described as 2/200,000. That such cosmetic variations in a statistic can affect its perceived value suggests that the party charged with conveying data has substantial power.

Finally, we note that jurors responded similarly to the evidence when there were 1 and 2 exemplars. This result hints that the absolute number of available exemplars may be less important than whether such exemplars are called to mind at all.

GENERAL DISCUSSION

We identified and tested an exemplar-cuing theory for how people assign weight to low-probability statistical evidence. According to this theory, people become more sensitive to low-probability events when descriptions of those events make exemplars more readily available. Exemplars can be triggered in different ways. In Experiment 1, we focused on the interactive roles played by form (probability or frequency), target (single or multiple), and reference class in producing or suppressing exemplars. In Experiment 2, we controlled exemplar access by varying the numerator and denominator of the incidence rate while holding the ratio constant. In both experiments, we found that people were less persuaded by low-probability evidence (a DNA match between a suspect and blood recovered from a crime scene) when it was presented in an exemplar-conducive way than when it was not. We have argued that this result occurred because the exemplars promoted thoughts about other individuals whose DNA might also match the evidence.

The Normative Issue

Much as availability is often a reasonable cue to frequency (Tversky & Kahneman, 1973), the availability of exemplars is sometimes a reasonable cue to diagnosticity. As incidence rates become small, exemplars become less available and concern that the low-probability event will occur on any single trial diminishes. However, there is no normative justification for exemplar effects that arise through strategic selection of statistical form, target, or reference class. When exemplars are manipulated in this way, the normative status of exemplar-based reasoning is suspect. Therefore, we make no representations about the accuracy of judgments that are influenced by the absence or presence of exemplars.

Practical Implications

A legal implication of our findings is that prosecutors and their experts may wish to present low-probability evidence with single targets, perhaps with incidence-rate numerators less than 1. Defense attorneys and their experts may wish to use multiple targets with incidence-rate numerators greater than or equal to 1. Our data may also be of interest to courts, scientific groups, and legislative bodies that develop standards for the admissibility of scientific and statistical evidence in court (cf. National Research Council, 1996). For example, it might be appropriate to present statistical evidence to jurors in multiple ways to minimize the influence of any particular bias.

The data have practical implications beyond the legal arena. Seat belts may seem like a better idea when the risk of dying in a car accident is communicated with readily available exemplars than when exemplars are not called to mind (cf. Slovic, Fischhoff, & Lichtenstein, 1978). The attractiveness of raffle tickets may be increased by presenting the chance of winning in multiple-target form using a large reference population. Genetic counselors may influence their patients' fears about diseases and vaccines by describing the relevant incidence rates with unconventional numerators. Of course, such strategic manipulation raises ethical issues (Slovic, 1999).

Acknowledgments—This research was supported by a grant from the National Science Foundation to the first author (SBR 9819305). We thank Brian Gibbs and Molly Mercer for helpful comments.

REFERENCES

Brase, G.L., Cosmides, L., & Tooby, J. (1998). Individuating, counting and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, *127*, 3–21.

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.

Gastwirth, J.L. 2000. *Statistical science in the courtroom*. New York: Springer.

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

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communication of statistical information. *Science*, *290*, 2261–2262.

[5]In another study (Koehler, 2002), we found that mock jurors who were provided with even larger ratios–0/10,000 and 10,000/10,000,000 for the .001 incidence rate and 10/1,000,000 and 10,000/1,000,000,000 for the .00001 incidence rate—responded to the evidence just as did jurors who were provided with the equivalent ratios that contained 1 in the numerator (i.e., 1/1,000 and 1/100,000, respectively). Inspection of jurors' response sheets indicated that many jurors simplified the large ratios to smaller ratios by crossing out zeroes. Apparently, ratios that are hard to understand but easy to simplify are simplified before being processed.

Jones, S.K., Jones, K.T., & Frisch, D. (1995). Biases of probability assessment: A comparison of frequency and single-case judgments. *Organizational Behavior and Human Decision Processes*, *61*, 109–122.

Kahneman, D., & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *92*, 136–153.

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.

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. (2001). When are people persuaded by DNA match statistics? *Law and Human Behavior*, *25*, 493–513.

Koehler, J.J. (2002). [Jurors' decision making when incidence ratios are expressed with large numerators and denominators]. Unpublished raw data.

Macchi, L. (1995). Pragmatic aspects of the base rate fallacy. *The Quarterly Journal of Experimental Psychology*, *48*, 188–207.

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: Beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes*, *82*, 217–236.

Miller, D.T., Turnbull, W., & McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology*, *57*, 581–589.

National Research Council, Committee on DNA Forensic Science. (1996). *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basis beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*, 972–987.

Slovic, P. (1999). Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis*, *19*, 689–701.

Slovic, P., Fischhoff, B., & Lichtenstein, P. (1978). Accident probabilities and seat belt usage: A psychological perspective. *Accident Analysis and Prevention*, *10*, 281–285.

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.

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.

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

Windschitl, P.D., & Wells, G.L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, *2*, 343–364.

(Received 6/11/02; Revision accepted 6/16/03)