Familiarity Bias and Belief Reversal in Relative Likelihood Judgment

Craig R. Fox and Jonathan Levav

Fuqua School of Business, Duke University

People are often called on to make an assessment of the relative likelihood of events (e.g., which of two investments is more likely to outperform the market?) and their complements (which of the two investments is more likely to perform no better than the market?). Probability theory assumes that belief orderings over events and their complements should mirror each other (i.e., $P(A) \geq P(B)$ iff $P(not\text{-}A) \leq P(not\text{-}B)$). This principle is violated in several surveys in which we asked people to assess the relative likelihood of familiar versus unfamiliar events. In particular, respondents are biased to view more familiar events (and their complements) as *more likely* than less familiar events (and their complements). Similarly, we observe that subjects are biased to view less familiar events (and their complements) as *less likely* than more familiar events (and their complements). Further studies demonstrate that the familiarity bias is less pronounced among subjects who are asked to judge the *probability* of each event rather than which event is *more likely*. Moreover, a greater proportion of subjects rate the more familiar event as *more likely* than assign a higher probability to that event. These patterns can be construed as *belief reversals*, analogous to the preference reversal phenomenon in decision making. The data are consistent with a contingent weighting model in which the process of judging relative likelihood biases attention toward evidence supporting the target hypothesis (and away from evidence supporting its complement). Because it is easier to recruit evidence supporting

The authors are grateful to Jim Bettman, David Budescu, Greg Fischer, Chip Heath, John Payne, and Michel Regenwetter for helpful comments and suggestions. The present work also benefited from discussions with Lyle Brenner and Yuval Rottenstreich who have run studies similar to problems 4 and 5, comparing "more likely" and "less likely" elicitation modes. We thank Richard Birke and James Bonk for assistance collecting data.

Address correspondence and reprint requests to Craig R. Fox, Fuqua School of Business, Duke University Box 90120, Durham, NC 27708. E-mail: cfox@mail.duke.edu.

Published in: ***Organizational Behavior and Human Decision Processes***

familiar events than unfamiliar events, this skewed attention causes both familiar events and their complements to be judged more likely, on average, than unfamiliar events and their complements.

Key Words: belief reversal; familiarity bias; support theory; contingent weighting; judgment under uncertainty.

People are frequently called on to order the strength of their beliefs over events. For instance, in deciding which car to purchase, a consumer may ask himself which model is more likely to break down; in deciding among various treatments for a particular disease, a patient might assess which treatment is more likely to succeed; in choosing among mutual funds, an investor might contemplate which fund is more likely to outperform the S&P 500 index. Curiously, although the last 30 years have witnessed an explosion of interest in the psychological processes underlying cardinal judgment of probability (see e.g., Kahneman, Slovic, & Tversky, 1982) and verbal expression of belief strength (Budescu & Wallsten, 1995; see also Fox & Irwin, 1998), the psychological process underlying ordinal judgment of belief has been largely ignored.

With the notable exception of studies comparing probability and frequency judgment (e.g., Fiedler, 1988; Gigerenzer & Hoffrage, 1995; Tversky & Kahneman, 1983), researchers have typically taken for granted that the mode by which beliefs are elicited has no effect on their ordering. In particular, they assume that the explicit comparison of two events (e.g., "which event do you think is more likely?") yields the same ordering as the consecutive cardinal evaluation of each of those events ("what is the probability of each event?"). This assumption is surprising in light of the fact that numerous studies have shown that measured preferences are often influenced by the specific way in which they are elicited (Payne, 1982; Payne, Bettman, & Johnson, 1993; Slovic & Lichtenstein, 1983). The present investigation is prompted by the conjecture that elicitation mode can also affect the ordering of beliefs over events.

Consider the case of preferences. In one study, participants judged the relative attractiveness of a prospect that offered a .31 chance of receiving $16 and a second that offered a .97 chance of receiving $4. Most people priced the first prospect higher than the second, yet indicated a preference to receive the second rather than the first (Tversky, Slovic, & Kahneman, 1990). Such judgment–choice preference reversals can be explained by the *compatibility principle*: the weight that a particular feature of a stimulus receives is enhanced by its compatibility with the response mode (Tversky, Sattath, & Slovic, 1988; for alternative accounts see Goldstein & Einhorn, 1987; Mellers, Ordóñez, & Birnbaum, 1992). The rationale for this principle is that characteristics of the elicitation task prime the most compatible features of the stimulus and that noncompatibility between input and output increases effort and error which reduces confidence and impact (Fitts & Seeger, 1953; Wickens, 1984). In the example above, the pricing task enhances the weight afforded the dollar amount of each prize so that the low-probability prospect (.31; $16) is priced higher, whereas the choice task enhances the weight afforded the probability of winning so that the prospect with smaller payoff, but higher probability (.97; $4), is more often chosen (Tversky, Slovic, & Kahneman, 1990). Similarly, studies of multiattribute choice suggest that positive attributes are weighted more heavily relative to negative attributes when choosing rather than rejecting options. If one option has both more positive and more negative attributes than another option, situations may arise in which most people choose the "enriched" option over the "impoverished" option *and* most people reject the enriched option in favor of the impoverished option (Shafir, 1993).

Tversky, Sattath, and Slovic (1988) assert that compatibility can affect choices not only through the *formal* correspondence between explicitly stated features (e.g., probability and prize amount) and the response mode (e.g., pricing vs choice), but also through the *semantic* correspondence between subjective features and the response mode. That is, compatibility may operate on features that are not explicitly provided, but rather are spontaneously recruited by subjects. For instance, Tversky's contrast model (1977) represents the similarity of two objects as a linear combination of their perceived common and distinctive features. The compatibility hypothesis suggests that common features loom larger in judgments of similarity than dissimilarity and that distinctive features loom larger in judgments of dissimilarity than similarity. Hence, a pair of objects with several common features and several distinctive features might be judged both more "similar" and more "dissimilar" than a pair of objects with fewer common and fewer distinctive features. Tversky and Gati (1978) report a study in which most subjects in one group judged a pair of familiar items (East and West Germany) as more *similar* than a pair of unfamiliar items (Ceylon and Nepal), but most subjects in a second group judged the familiar items as more *dissimilar* than the unfamiliar items.

What relevant features might be spontaneously recruited by people when assessing probabilistic beliefs? One tradition in the literature, beginning with Keynes (1921) and continuing more recently in support theory (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994), interprets judged probability as quantification of the *balance* of evidence favoring a proposition relative to evidence opposing that proposition. This balance of evidence may be supported by a large or small total *mass* of evidence. Hence, a financial analyst may believe that there are many equally compelling arguments in favor and against a rise in U.S. unemployment next quarter and therefore estimate the probability of an increase to be 50%. That same analyst might also recruit a vague argument or two both favoring and opposing a rise in Djibouti's unemployment rate, again yielding a forecast of 50%. In this case the perceived support for the highly familiar hypothesis, H (U.S. unemployment will rise), is strong, balanced by strong perceived support for its complement, $\overline{H}$ (U.S. unemployment will drop or remain unchanged); the perceived support for the less familiar hypothesis, L (Djibouti unemployment will rise), is weak, balanced by weak perceived support for its complement, $\overline{L}$ (Djibouti unemployment will drop or remain unchanged). Thus, the *balance* of evidence—and therefore the judged probability—is approximately the same in both cases, but this balance is supported by a larger total *mass* of evidence in the former case.[1]

If people typically consider evidence both for and against hypotheses when assessing belief strength, how might the relative weighting of this evidence be affected by the elicitation mode? Fischer, Carmon, Ariely, and Zauberman (1999) argue that people tend to give greater weight to more salient attributes when the goal of the task is to differentiate between objects. In particular, they assert that choice and strength of preference (which require the decision maker to differentiate between options) promote greater weight to the "prominent" attribute, whereas pricing tasks (which require the decision maker to match an option with a sure amount of money) promote more equal weighting. Analogously, we suggest that when a person is asked to differentiate which hypothesis, H or L, is "more likely," evidence supporting these focal hypotheses is more salient and therefore receives more weight. It may not be necessary to make a more thorough assessment of evidence that can be recruited for the two corresponding complementary hypotheses, $\overline{H}$ and $\overline{L}$. For example, if I am asked to evaluate whether it is *more likely* to rain tomorrow in Phoenix or Seattle, I may compare how easy it is to recall rainy days this time of year in each city, but I might not consider how easy it is to recall days without rain. In contrast, the probability scale requires a mapping of absolute degree of belief onto the unit interval which necessitates consideration of complementary hypotheses. If the perceived balance of evidence entirely favors the target hypothesis, this belief is mapped into the number 1; if the perceived balance of evidence entirely favors the complementary hypothesis, this belief is mapped into the number 0; if the perceived balance of evidence is equal, this belief is mapped into the number .5.[2] Hence, if I am asked to evaluate the *probability* of rain tomorrow in Phoenix and the *probability* of rain tomorrow in Seattle, I must compare how easy it is to recall rainy days this time of year to how easy it is to recall days without rain, separately for each city.

This notion that people give more weight to the focal hypothesis when judging relative likelihood is also consistent with a well-established body of research suggesting that people typically pursue simplifying strategies when making complex judgments or choices in order to overcome limitations in their information processing capacity (Fiske & Taylor, 1991; Kahneman, Slovic, & Tversky, 1982, 2000; Payne, Bettman, & Johnson, 1993). Note that the view of subjective probability that we have cited holds that in order to judge the likelihood of a

[1] The relationship between balance of evidence, mass of evidence, and calibration is discussed by Griffin and Tversky (1992), who argue that overconfidence occurs when the balance of evidence strongly favors one hypothesis but the mass of evidence supporting that balance is low, whereas underconfidence occurs when the balance of evidence is more even but the mass of evidence supporting that balance is high.

[2] In this respect, it is worth noting that research on verbal probabilities reveals that people agree most in their numerical interpretation of terms at the endpoints of the probability scale (e.g., *impossible, certain*) in which all evidence opposes or favors the focal hypothesis and the middle of the scale (e.g., *even odds*) in which there is equal evidence (see Budescu & Wallsten, 1995).

hypothesis, H, a person must weigh evidence for the focal hypothesis (H) against evidence for its alternative ($\overline{H}$). To evaluate the *relative* likelihood of two distinct hypotheses, H and L, however, a person would be required not only to evaluate the balance of evidence for H versus $\overline{H}$ and the balance of evidence for L versus $\overline{L}$, but also to render a second-order comparison of these judgments. The much simpler task of comparing evidence for H versus L gives the same result under most circumstances.

For these reasons, we conjecture that evidence for alternative hypotheses ($\overline{H}$ and $\overline{L}$) will loom larger in cardinal judgment of probability than in ordinal judgment of relative likelihood. It seems reasonable to establish a belief ordering by comparing the evidence that one can recruit for each of these hypotheses (H, L). However, it seems impossible to qualify or quantify one's absolute degree of belief that a particular hypothesis will obtain without at least some consideration of the balance of evidence favoring versus opposing that hypothesis (H vs $\overline{H}$, L vs $\overline{L}$). We now proceed to formalize this conjecture and generate testable implications.

CONTINGENT WEIGHTING, FAMILIARITY BIAS, AND BELIEF REVERSAL

We begin with the theoretical foundation of support theory (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994), in which subjective probability is not attached to events, as it is in other models, but rather to descriptions of events, called *hypotheses*. Hence, two descriptions of the same event may be assigned different probabilities (i.e., the model is nonextensional).[3] Support theory assumes that each hypothesis A has a nonnegative support value $s(A)$ corresponding to the strength of evidence for this hypothesis. The judged probability $P(A, \overline{A})$ that hypothesis A rather than $\overline{A}$ holds, assuming that one and only one of them obtains is given by:

$$P(A, \overline{A}) = \frac{s(A)}{s(A) + s(\overline{A})}. \tag{1}$$

Thus, judged probability is interpreted as support for the focal hypothesis A relative to the alternative hypothesis $\overline{A}$. For example, the probability of rain tomorrow (A) rather than no rain ($\overline{A}$) is assumed to be the support for rain divided by the sum of support both for and against rain. It is convenient to translate Eq. (1) into an odds metric:

$$R(A, \overline{A}) \equiv \frac{P(A, \overline{A})}{1 - P(A, \overline{A})} = \frac{s(A)}{s(\overline{A})}. \tag{2}$$

Note that R is a notational device that is derived from judgments of probability,

[3] In this paper we will assume a canonical description of each event and will therefore not distinguish between events and hypotheses.

and that the ordering of hypotheses by odds is formally equivalent to the ordering of hypotheses by judged probability.

Contingent Weighting in Ordering Beliefs

Consider two (not necessarily exclusive) hypotheses H and L whose complements are $\overline{H}$ and $\overline{L}$, respectively. Let $\geq_i$ be the belief ordering of hypotheses under elicitation mode i ($i = P, \mu$), where P refers to the belief ordering inferred from separately evaluated judged probabilities and μ refers to direct assessment of which hypothesis is "more likely." It readily follows from Eq. (2) that $R(H, \overline{H}) \geq R(L, \overline{L})$ iff $\frac{s(H)}{s(\overline{H})} \geq \frac{s(L)}{s(\overline{L})}$, so that

$$H \geq_p L \text{ iff } \log s(H) - \log s(\overline{H}) \geq \log s(L) - \log s(\overline{L}).$$

This is merely a special case of the contingent weighting model (Tversky, Sattath, & Slovic, 1988):

$$H \geq_i L \text{ iff } \alpha_i \log s(H) - \beta_i \log s(\overline{H}) \geq \alpha_i \log s(L) - \beta_i \log s(\overline{L}), \qquad (3)$$

with $i = P$, $\alpha_p = 1$, and $\beta_p = 1$. Here α_i and β_i reflect the relative weight in response mode i of evidence favoring the focal and alternative hypotheses, respectively. In support theory the focal and alternative hypotheses receive equal and opposite weight. We speculate that the alternative hypothesis receives less weight than the focal hypothesis in relative likelihood judgment: $\beta_\mu/\alpha_\mu \leq 1$. Our conjecture that the alternative hypotheses will loom larger in cardinal judgment of probability than in ordinal judgment of relative likelihood can be expressed as $\beta_p/\alpha_p \geq \beta_\mu/\alpha_\mu$.

Familiarity Bias and Belief Reversal

Note that if $\alpha_\mu >> \beta_\mu$, then belief ordering is essentially determined by support for the focal hypothesis. Note also that the belief ordering over H and L will be the same regardless of the relative weight to the focal versus alternative hypothesis whenever the events are equally familiar so that the total amount of support for the focal and alternative hypotheses is the same for both events (i.e., $s(H) + s(\overline{H}) = s(L) + s(\overline{L})$).[4] However, if one event is more familiar than another, so that support both for and against one hypothesis is greater than support for and against a second hypothesis ($s(H) > s(L)$ and $s(\overline{H}) > s(\overline{L})$), then two interesting patterns emerge. First, situations can arise in which the familiar event is deemed both more likely to occur than the unfamiliar event ($H >_\mu L$) and more likely *not* to occur than the unfamiliar event ($\overline{H} >_\mu \overline{L}$).

[4] To see why, suppose the total amount of support for both pairs of hypotheses equals some constant (i.e., $s(H) + s(\overline{H}) = s(L) + s(\overline{L}) = C$). In this case, $s(H) \geq s(L)$ iff $s(\overline{H}) \leq s(\overline{L})$ so that $H \geq_i L$ for all $\alpha > 0$ and $\beta \geq 0$.

Second, situations can arise in which the more familiar event is deemed more likely ($H >_{\mu} L$) but is assigned a lower probability ($L >_{p} H$).

These patterns cannot be reconciled with prevailing models of subjective probability. First, probability theory and most descriptive models of judgment under uncertainty implicitly assume that the method by which beliefs are elicited has no effect on their ordering. Hence,

$$H >_i L \text{ iff } H >_j L$$

for all elicitation modes i and j that are normatively equivalent. For example, Harold is judged "nonlikely" to pass the exam than Larry if and only if Harold is assigned a higher probability of passing than Larry. Second, probability theory and many descriptive models predict a reflection in the ordering of beliefs over events versus the ordering of their complements:

$$H >_i L \text{ iff } \overline{L} >_i \overline{H}.$$

That is, H is deemed more likely to occur than L if and only if the complement of L (i.e., L does not occur) is deemed more likely than the complement of H (i.e., H does not occur).[5] For example, Harold is more likely to *pass* the exam than Larry if and only if Larry is more likely to *fail* the exam than Harold.

The empirical section of the paper is organized as follows. The first set of studies provides evidence for a *familiarity bias* when people explicitly order their beliefs by indicating which hypothesis they consider to be "more likely." A second set of studies replicates this phenomenon among people who are asked to order their beliefs by which hypothesis they consider to be "less likely." Finally, a third set of studies compares the magnitude of the familiarity bias in ordinal (more likely) versus cardinal (probability) judgment and documents reversals in belief orderings across these two modes.

ORDERING BY WHICH EVENT IS "MORE LIKELY"

Let $\Pi(H >_i L)$ be the proportion of respondents who order their belief in hypothesis H above their belief in hypothesis L, using response mode i. Probability theory and support theory predict that response proportions should be equal, whether one orders belief in these hypotheses or belief in their complements:

$$\Pi(H >_P L) = \Pi(\overline{L} >_P \overline{H}).$$

The notion that the focal hypothesis looms larger than the alternative hypothesis in relative likelihood judgment, however, implies a different pattern. Under

[5] A sufficient condition for this prediction is that the measure of belief strength for an event and its complement sum to a constant (that may or may not be one). Necessary conditions are considerably weaker. See Fox and Levav (2000).

this condition, if subjects can recruit more support both for and against one hypothesis than they can recruit for and against another (i.e., $s(H) \geq s(L)$ and $s(\overline{H}) \geq s(\overline{L})$), then the proportion of subjects judging the more familiar hypothesis *more likely* will be greater than the proportion of subjects judging the complement of the less familiar hypothesis *more likely*:

$$\Pi(H >_{\mu} L) \geq \Pi(\overline{L} >_{\mu} \overline{H}). \tag{4}$$

(Sufficient conditions for Eq. (4) are actually less restrictive with respect to support. See Appendix 1.)

All problems that follow were presented to participants as pencil-and-paper questionnaires. In each problem respondents were asked to judge the relative likelihood of a familiar event (H or $\overline{H}$) and an unfamiliar event (L or $\overline{L}$): one group was asked to compare H and L, and one group was asked to compare $\overline{H}$ and $\overline{L}$. The order in which items were presented was counterbalanced in all studies.

To test the familiarity bias hypothesis (Eq. (4)), we asked Duke University business students to complete the following item concerning upcoming sporting events against the University of North Carolina (UNC). Students were entered in a drawing for a $20 prize in exchange for their participation.

Problem 1: Duke University Sports (N = 135). Which of the following two events do you think is more likely to occur (please check one):

	$>_{\mu}$	
_____ Duke men's basketball beats UNC men's basketball at Duke's Cameron Indoor Stadium in January 1999.	(H) 75%	*Group 1*
_____ Duke men's fencing beats UNC men's fencing at Duke's Card Gym in January 1999.	(L) 25%	
* * * * *		
_____ UNC men's basketball beats Duke men's basketball at Duke's Cameron Indoor Stadium in January 1999.	($\overline{H}$) 44%	*Group 2*
_____ UNC men's fencing beats Duke men's fencing at Duke's Card Gym in January 1999.	($\overline{L}$) 56%	

Duke students spend a great deal of time following men's basketball, and UNC is Duke's cross-town rival. However, most Duke students know very little about men's fencing. We therefore assumed that questions concerning basketball (labeled H, $\overline{H}$) would be highly familiar, and questions concerning fencing (labeled L, $\overline{L}$) would be less familiar. The number listed beside each of these labels is the proportion of subjects in a given group who indicated that the corresponding event was *more likely* to occur than the event with which it was paired. Probability theory and support theory predict that the proportion of subjects joined by the arrowheads should be equal, whereas the present account predicts that the proportions will be greater for H and $\overline{H}$ than for $\overline{L}$ and L, respectively.

Results were consistent with the familiarity bias hypothesis: $\Pi(H >_{\mu} L) = .75 \geq .56 = \Pi(\overline{L} >_{\mu} \overline{H})$, $z = 2.37$, $p < .01$, one-tailed. The proportion of students who said that a Duke victory was more likely in basketball than in fencing was larger than the proportion of students who said a UNC victory was more likely in fencing than in basketball. To see why this pattern represents a bias, note that most students (75%) said that a Duke victory was more likely in basketball than fencing. Therefore, probability theory would predict that a minority of students (25%) should indicate that a UNC victory (i.e., Duke loss) is more likely in basketball than fencing. However, nearly *half* of the students in our sample (44%) indicated such a belief.

To replicate this item we presented business students with a familiar and an unfamiliar mutual fund. Every respondent (100%) in a sample of weekend MBA students ($N = 93$) reported that they were more familiar with Fidelity's Magellan fund than Lord Abbott's Affiliated fund. We next approached a separate sample of Duke daytime and weekend MBA students and asked them to make judgments concerning the future performance of these funds.

Problem 2: Mutual Funds (N = 228). Which *of the following two events do you think is more likely to occur (please check* one):

		$>_{\mu}$
_____ Fidelity's Magellan Fund **underperforms** the S&P 500 Index over the next 12 months.	(H)	38%
_____ Lord Abbott's Affiliated Fund **underperforms** the S&P 500 Index over the next 12 months.	(L)	62%

* * * * *

_____ Fidelity's Magellan Fund performs at **least as well as** the S&P 500 Index over the next 12 months.	($\overline{H}$)	73%
_____ Lord Abbott's Affiliated Fund performs at **least as well as** the S&P 500 Index over the next 12 months.	($\overline{L}$)	27%

There were no significant differences between responses of the weekend and daytime MBA students, so their data were combined. As predicted, respondents exhibited a bias in favor of the more familiar fund: $\Pi(H >_{\mu} L) = .38 \geq .27 = \Pi(\overline{L} >_{\mu} \overline{H})$, $z = 1.79$, $p < .05$. A larger proportion of subjects said the familiar fund was more likely to perform well, compared to the proportion of subjects who said that the unfamiliar fund was more likely to perform poorly.

The foregoing examples demonstrate a bias in favor of familiar hypotheses and their complements. We have interpreted this finding as a consequence of evidence for the focal hypothesis looming larger than evidence for the alternative hypothesis. We assume that evidence for and against the focal hypothesis is spontaneously recruited by respondents and that they recruit more abundant evidence both for and against the focal hypothesis in the case of the more familiar event. The next study attempts to experimentally manipulate the abundance of evidence both for and against the target hypothesis. We created profiles of two employees suspected of a theft: one employee (employee A) is associated with three incriminating and three exculpatory pieces of evidence, and another employee (employee B) is associated with six pieces of evidence that are neither incriminating nor exculpatory.

> *Problem 3: Corporate Theft ($N = 144$).* Imagine that you head a department in your company. One afternoon, during lunch hour, a laptop computer belonging to the department disappears from a desk in an employee's office. You decide to launch an inquiry. In the course of your investigation you interview two employees.
>
> *Which of the following two employees do you think is more likely to be [innocent/ guilty] of the theft?*

Employee A

- Has consistently earned extremely high performance appraisals from supervisors.
- Has access to the master key to all the offices.
- Another employee claims to have been at a restaurant with him for most of the lunch hour.
- Was dismissed from his previous job on suspicion of theft, but insists he was framed.
- Already owns a laptop computer.
- Has a history of bitter disagreements with the person from whom the computer was stolen.

Employee B

- Is often one of the last employees to leave the office at the end of the day.
- Has been working at the firm for many years.
- Has been known to spend his lunch hours at a cafeteria in an adjacent building.
- Has previously remarked how impressed he is with the computer equipment provided by your company.
- Is well-liked by most (but not all) employees.
- Has expressed interest in learning more about the World Wide Web.

	$>_\mu$		$>_\mu$
Employee A *innocent* $= H$	41%	Employee A *guilty* $= \overline{H}$	79%
Employee B *innocent* $= L$	59%	Employee B *guilty* $= \overline{L}$	21%

As a manipulation check, we asked a separate sample of business students ($N = 56$) to evaluate how guilty or innocent each fact (listed in a random order) would make an employee appear, scored on a −5 (extremely guilty) to 0 (neutral) to +5 (extremely innocent) scale. The median response to each of the facts for the neutral employee (listed here as employee B) was 0, and the median response to each of the facts for the mixed employee (listed here as employee A) were +1, +2, +1 for the exculpatory facts (first, third, and fifth items listed, respectively) and −1, −3, −1 for the incriminating facts (second, fourth, and sixth items listed, respectively). When the absolute value of these scores were summed, 93% of respondents reported a higher total for the facts associated with the mixed employee than the facts associated with the neutral employee.

Problem 3 was presented to business students at Duke University as part of a questionnaire packet that also included unrelated items. Respondents were compensated for an hour's participation with a $10 donation to a charity. Results conformed to our prediction. Subjects indicated that the mixed employee (A) was more likely to be innocent nearly twice as often as they said that the neutral employee (B) was more likely to be guilty: $\Pi(H >_{\mu} L) = .41 \geq .21 = \Pi(\overline{L} >_{\mu} \overline{H})$, $z = 2.65$, $p < .005$. Apparently, focusing respondents' attention on guilt versus innocence can have a dramatic impact on their impressions of suspects. Moreover, we successfully replicated the "familiarity" bias by experimentally manipulating the mass of evidence favoring both a hypothesis and its complement rather than assuming that subjects spontaneously recruit more evidence in more familiar domains.

ORDERING BY WHICH EVENT IS "LESS LIKELY"

Thus far we have described problems in which subjects were asked to order their beliefs over events by indicating which they believed was *more likely* to occur. Consistent with our prediction, respondents were biased to regard the more familiar event (or the event associated with stronger evidence) as more likely. Should we expect a different pattern if we were to ask subjects to order their beliefs by which event they think is *less likely*? Previous research on compatibility effects has documented cases in which reversing the polarity of elicitation (e.g., from choosing to rejecting) alters the relative weighting of attributes (negative attributes receive more weight relative to positive attributes in rejecting compared to choosing), leading to a reversal of judgment or preference (Shafir, 1993; Tversky & Gati, 1978). For instance, Shafir (1993) found that for some items respondents were more likely both to *choose* and *reject* an option characterized by several positive and negative attributes over an option characterized by neutral attributes. In contrast, the present account suggests that relative likelihood judgment, regardless of polarity, promotes greater attention to the designated focal hypotheses. Our intuition is that "more likely" prompts an evaluation of which focal hypothesis is *more* easy to imagine and that "less likely" prompts an evaluation of which focal hypothesis is *less* easy to imagine—with equal neglect in both cases to the corresponding complementary hypotheses. Hence, respondents should be biased to view the less familiar hypothesis and its complement as *less likely* than the more familiar hypothesis and its complement. Thus, whereas previous applications of the compatibility principle predict a reversal in responses when the polarity of the elicitation mode is reversed, we predict that reversal of polarity will preserve belief orderings. In other words, just as subjects will be biased to rate more familiar events *more likely*, they will be biased to rate less familiar events *less likely*.

More formally, in Eq. (3), if we let $H \geq_\lambda L$ refer to the ordering of H over L in *less likely* mode, then we propose that $\beta_\lambda/\alpha_\lambda \approx \beta_\mu/\alpha_\mu$. Note that $H \geq_\lambda L$ ("L is less likely than H") expresses the same belief ordering as $H \geq_\mu L$ ("H is more likely than L"). The familiarity bias prediction for the *less likely* elicitation mode can be formalized as $\Pi(H >_\lambda L) \geq \Pi(\overline{L} >_\lambda \overline{H})$. In extreme cases, most people might say that the less familiar event is less likely to occur *and* less likely not to occur. We also expect that the proportion of respondents judging L less likely than H will be about the same as the proportion of respondents judging H more likely than L, $\Pi(H >_\lambda L) \approx \Pi(H >_\mu L)$. Note that if the alternative hypothesis loomed larger in "less likely" judgment than in "more likely" judgment, as would be predicted by an account analogous to Tversky & Gati (1978) and Shafir (1993), then we would instead observe $\Pi(H >_\lambda L) \leq \Pi(H >_\mu L)$.

To test these hypotheses, we recruited undergraduate students in an introductory psychology class and a campus walkway at Duke University. The same procedure was used as in Problems 1–3; in addition, two conditions were included in which respondents were asked to judge which of the two events is *less likely*.

Problem 4: Temperatures (N = 333). We would like to know what you think the average high temperature will be next week in different U.S. cities. *Please circle the event that you think is* ***[more likely/less likely]*** *to occur*:

		$>_\mu$	$>_\lambda$
Next week's average high temperature in Durham, NC will be **below** 56 degrees Fahrenheit.	(H)	50%	50%
Next week's average high temperature in Chandler, OK will be **below** 56 degrees Fahrenheit.	(L)	50%	50%
* * * * *			
Next week's average high temperature in Durham, NC will be **at least** 56 degrees Fahrenheit.	$(\overline{H})$	63%	62%
Next week's average high temperature in Chandler, OK will be **at least** 56 degrees Fahrenheit.	$(\overline{L})$	37%	38%

We presumed that students would be more familiar with temperature in Durham, NC (where Duke University is located) than temperature in Chandler, OK. Because there were no significant differences between subject populations, the data were pooled. Results again conformed to our prediction. Respondents exhibited a bias in favor of the familiar hypothesis and its complement when judging which event is *more likely*, $\Pi(H >_{\mu} L) = .50 \geq .37 = \Pi(\overline{L} >_{\mu} \overline{H})$, $z = 1.70$, $p < .05$, and an identical effect when judging which event is *less likely*, $\Pi(H >_{\lambda} L) = .50 \geq .38 = \Pi(\overline{L} >_{\lambda} \overline{H})$, $z = 1.58$, $p < .06$. Stated differently, although half the students thought that it was more likely to be cold (less than 56 degrees) in Durham than in Chandler, nearly two thirds (63%) thought it was more likely to be warm (at least 56 degrees) in Durham than in Chandler. Similarly, although half the students thought that it was less likely to be cold in Chandler than in Durham, nearly two thirds (62%) thought that it was less likely to be warm in Chandler than in Durham. As expected, there was no difference between *more likely* and *less likely* elicitation modes ($\Pi(H >_{\mu} L) - \Pi(H >_{\lambda} L) = 0.00$, $z = 0.00$, n.s.; $\Pi(\overline{H} >_{\mu} \overline{L}) - \Pi(\overline{H} >_{\lambda} \overline{L}) = .01$, $z = 0.13$, n.s.), and the combined (experimentwise) familiarity bias was statistically significant ($z = 2.22$, $p < .05$).

We replicated this effect by recruiting customers of a video store during the weekend preceding the 1998 Academy Awards ceremony and asking them to order their beliefs over potential award winners. Surveys were left in a box at the checkout counter, and respondents were told that one participant would be selected at random to receive 10 free video rentals.

Problem 5: Academy Awards (N = 116).

• The following films have been nominated for the 1998 **best picture** academy award: "As Good As It Gets," "The Full Monty," "Good Will Hunting," "L.A. Confidential," and "Titanic."

• The following films have been nominated for the 1998 **best foreign language film** academy award: "Beyond Silence," "Character," "Four Days in September," "Secrets of the Heart," and "The Thief."

Which of the following two options, A or B, do you think is ***[more likely/ less likely]*** *to occur (please check one of the following two options)*:

		$>_{\mu}$	$>_{\lambda}$
____Option A: **"The Full Monty"** OR **"Titanic"** wins best picture.	(H)	76%	80%
____Option B: **"Character"** OR **"The Thief"** wins best foreign language film.	(L)	24%	20%

* * * * *

____Option A′: **"As Good As It Gets"** OR **"Good Will Hunting"** OR **"L.A. Confidential"** wins best picture.	($\overline{H}$)	48%	57%
____Option B′: **"Beyond Silence"** OR **"Four Days in September"** OR **"Secrets of the Heart"** wins foreign language film.	($\overline{L}$)	52%	43%

Results again accord with our prediction. Respondents exhibited a bias in favor of the familiar hypothesis and its complement when judging which event is *more likely* ($\Pi(H >_{\mu} L) = .76 \geq .52 = \Pi(\overline{L} >_{\mu} \overline{H})$, $z = 1.94$, $p < .05$) and a similar pattern when judging which event is *less likely* ($\Pi(H >_{\lambda} L) = .80 \geq .43 = \Pi(\overline{L} >_{\lambda} \overline{H})$, $z = 3.19$, $p < .001$). Stated differently, although more than three quarters (76%) of respondents indicated that H (*Full Monty* or *Titanic* wins best picture) was more likely than L (*Character* or *The Thief* wins best foreign film), nearly half (48%) indicated that $\overline{H}$ (*As Good As It Gets*, *Good Will Hunting*, or *L.A. Confidential* wins best picture) was more likely than $\overline{L}$ (*Beyond Silence*, *Four Days in September*, or *Secrets of the Heart* wins best foreign film). Similarly, although more than three quarters (80%) of respondents indicated that L was less likely than H, more than half (57%) indicated that $\overline{L}$ was less likely than $\overline{H}$. As in the previous study, there was no significant difference between *more likely* and *less likely* elicitation modes (($\Pi(H >_{\mu} L) - \Pi(H >_{\lambda} L) = -0.04$, $z = -0.35$, n.s.; $\Pi(\overline{H} >_{\mu} \overline{L}) - \Pi(\overline{H} >_{\lambda} \overline{L})) = 0.09$, $z = 0.69$, n.s.), and the combined (experimentwise) familiarity bias was statistically significant ($z = 3.53$, $p < .001$). Thus, evidence from two studies strongly supports the notion that reversing polarity of the elicitation mode has no effect on the relative weighting of evidence, contrary to the prediction based on previous research (Tversky & Gati, 1978; Shafir, 1993).

CARDINAL VS ORDINAL JUDGMENT

The familiarity bias in relative likelihood judgment is motivated by the conjecture that support for the focal hypothesis looms larger than support for the alternative hypothesis when making ordinal comparisons between events. As noted earlier, this contrasts sharply with support theory, which assumes that focal and alternative evidence receive equal (and opposite) weight. It is easy to verify that Eq. (1) implies *binary complementarity*, $P(A, \overline{A}) + P(\overline{A}, A) = 1$, a property that has been found to hold reasonably well in numerous experimental studies (for reviews see Fox & Tversky, 1998; Tversky & Koehler, 1994; see also Fox, 1999; but see Brenner & Rottenstreich, 1999; Macchi, Osherson, & Krantz, 1999). This condition implies further that $\Pi(H >_{P} L) = \Pi(\overline{L} >_{P} \overline{H})$. Of course, binary complementarity may not hold perfectly when two events are juxtaposed. For instance, some respondents may judge the probability of the second event by anchoring on their judgment of the first, then adjusting according to whether they perceive the second to be more likely or less likely. Hence, rather than commit to a complete absence of bias for judged probability we predict that the familiarity bias will be less pronounced for judged probability than for judgments of which event is "more likely,"

$$\Pi(H >_{\mu} L) - \Pi(\overline{L} >_{\mu} \overline{H}) \geq \Pi(H >_{P} L) - \Pi(\overline{L} >_{P} \overline{H}). \tag{5}$$

Moreover, the notion that the focal hypothesis looms larger than the alternative hypothesis in relative likelihood judgment compared to probability judgment implies that the tendency to order high familiarity events over low familiarity events will be more pronounced for judgments of which event is "more likely" than for judgments of probability,

$$\Pi(H >_{\mu} L) \geq \Pi(H >_{P} L) \tag{6}$$

(See Appendix 1 for sufficient conditions.) Note that this tendency is expected to be stronger when $\Pi(H >_P L)$ is relatively small so that there are no ceiling effects. In fact, situations could arise in which most people judge the more familiar event to be more likely than the less familiar event but most people assign the more familiar event a lower probability. Such a pattern would constitute a *belief reversal*, akin to the preference reversal phenomenon observed in studies of choice. We begin by investigating the belief reversal phenomenon (Eq. (6)) and then proceed to describe studies in which we can also test for attenuation of the familiarity bias (Eq. (5)).

To test the belief reversal prediction we recruited first-year law students at Willamette University (located in Salem, Oregon). The students had spent a class session during the previous week discussing a case pending at the Oregon State Supreme Court and therefore were presumed to be highly familiar with the particulars of that case. By coincidence, we discovered that a case with similar facts and issues was pending in the Colorado State Supreme Court. We presumed that students were less familiar with this latter case.

> *Problem 6: Law Case* ($N = 117$). Please recall the Oregon Supreme Court arguments held here at the College of Law. The case of *State v. Smith* involved the question of whether a dog sniff constituted a search. That case is still pending. *People v. Reyes*, a case with similar facts and similar legal issues, is pending in the Colorado Supreme Court.

[More likely condition]

Which of the following is more likely (check one):

		$>_{\mu}$
_____ The APPELLANT (Smith) in *State v. Smith* prevails?	(H)	66%
_____ The APPELLANT (Reyes) in *People v. Reyes* prevails?	(L)	34%

[Probability condition]

		$>_{P}$
What is your best estimate of the probability of the APPELLANT (Smith) prevailing in State v. Smith?	(H)	49%
What is your best estimate of the probability of the APPELLANT (Reyes) prevailing in People v. Reyes?	(L)	51%

Results supported the belief reversal hypothesis (Eq. (6)): $\Pi(H >_{\mu} L) = .66 \geq .49 = \Pi(H >_P L)$, $z = 1.88$, $p < .05$. Although a roughly equal proportion of students indicated a higher probability for each of the appellants, the majority of students said they thought that it was *more likely* that the familiar appellant (Smith) would prevail.[6]

To test the hypothesis that the *probability* mode results in a less pronounced familiarity bias than the *more likely* mode (Eq. (5)), we presented the following item to members of the Duke University community at the Student Health Center's allergy clinic. Respondents were told that one participant would be selected at random to receive a $30 gift certificate for a local bookstore.

Problem 7: Politics ($N = 86$). [Which of the following two events do you think is **more likely** to occur (please check one):/Please indicate your best estimates of the probabilities of the following two events:]

		$>_\mu$	$>_P$
_______ The winner of the next U.S. Presidential election **is** member of the Democratic Party.	(H)	64%	36%
_______ The winner of the next British Prime Ministerial election **is** a member of the Labor Party.	(L)	36%	64%

* * * * *

_______ The winner of the next U.S. Presidential election **is not** a member of the Democratic Party.	$(\overline{H})$	76%	73%
_______ The winner of the next British Prime Ministerial election **is not** a member of the Labor Party.	$(\overline{L})$	24%	27%

Results support the attenuation hypothesis (Eq. (5)). We obtained a strong familiarity bias for *more likely* judgments, $\Pi(H >_\mu L) = .64 \geq .24 = \Pi(\overline{L} >_\mu \overline{H})$, $z = 2.89$, $p < .005$, but a nonsignificant effect for judged probability: $\Pi(H >_P L) = .36 \cong .27 = \Pi(\overline{L} >_P \overline{H})$, $z = 0.64$, n.s. The difference in these effects is statistically significant: $\Pi(H >_\mu L) - \Pi(\overline{L} >_\mu \overline{H}) = .40 > .09 = \Pi(H >_P L) - \Pi(\overline{L} >_P \overline{H})$, $z = 1.79$, $p < .05$. It is also worth pointing out that we replicated the belief reversal hypothesis (Eq. (6)) for *more likely* compared to *probability* modes in the Democrat vs Labor conditions ($\Pi(H >_\mu L) = .64 \geq .36 = \Pi(H >_P L)$, $z = 1.91$, $p < .05$) and found a nonsignificant tendency in the predicted direction for the not-Democrat versus not-Labor conditions ($\Pi(\overline{H} >_\mu \overline{L}) = .76 \geq .73 = \Pi(\overline{H} >_P \overline{L})$, $z = 0.23$, n.s.). On average, the experimentwise pattern of belief reversal is statistically significant[7] ($z = 2.12$, $p < .05$). It is worth

[6] The ordinal analysis of probabilities is a bit problematic as it allows for ties, whereas the *more likely* elicitation mode does not. In these analyses we break ties by assigning half of these subjects to the $H >_P L$ category and half to the $L >_P H$ category. The proportion of ties in the raw data are reported in Table 1.

[7] That is, the following index is significantly greater than zero:

$$\frac{[\Pi(H >_\mu L) - \Pi(H >_P L)] + [\Pi(\overline{H} >_\mu \overline{L}) - \Pi(\overline{H} >_P \overline{L})]}{2}.$$

pausing to emphasize our finding in the Democrat–Labor conditions: most people (64%) thought a Democrat was *more likely* to win than a Labor candidate, whereas most people (64%) assigned a higher probability to a Labor candidate winning than to a Democrat winning.

We attempted to replicate this result by recruiting undergraduates in an introductory chemistry class at Duke University on the eve of the Atlantic Coast Conference (ACC) men's basketball tournament.

> *Problem 8: College Basketball (N = 305).* [Which of the following two events do you think is **more likely** to occur (please check one):/Please indicate your best estimates of the probabilities (0–100%) of each of the following two events:]

		$>_\mu$	$>_P$
_____ Georgia Tech beats UNC in men's basketball tonight.	(H)	59%	24%
_____ Washington State beats Washington in men's basketball tomorrow night.	(L)	41%	76%

* * * * *

_____ UNC beats Georgia Tech in men's basketball tonight.	$(\overline{H})$	74%	64%
_____ Washington beats Washington State in men's basketball tomorrow night.	$(\overline{L})$	26%	36%

Georgia Tech and UNC are rivals of Duke in the ACC and were therefore presumed to be more familiar to Duke students than Washington State and Washington, which play in the Pacific-10 conference. Results of this survey were even stronger than those observed for Problem 7. We obtained a pronounced familiarity bias for *more likely* judgments, $\Pi(H >_\mu L) = .59 \geq .26 = \Pi(\overline{L} >_\mu \overline{H})$, $z = 4.41$, $p < .001$, and a tendency in the opposite direction for judged probability, $\Pi(H >_P L) = .24 \cong .36 = \Pi(\overline{L} >_P \overline{H})$, $z = -1.62$, n.s. by two-tailed test. Moreover, the interaction is highly significant, $\Pi(H >_\mu L) - \Pi(\overline{L} >_\mu \overline{H}) = .33 > -.12 = \Pi(H >_P L) - \Pi(\overline{L} >_P \overline{H})$, $z = 4.87$, $p < .001$, providing strong support for the attenuation hypothesis (Eq. (5)). Also, we replicated a belief reversal in the Georgia Tech winning versus Washington State winning conditions ($\Pi(H >_\mu L) = .59 \geq .24 = \Pi(H >_P L)$, $z = 4.69$, $p < .001$) and found a nonsignificant tendency in the predicted direction for UNC winning versus Washington winning ($\Pi(\overline{H} >_\mu \overline{L}) = .74 \geq .64 = \Pi(\overline{H} >_P \overline{L})$, $z = 1.35$, n.s.). On average, the experimentwise pattern is highly significant ($z = 4.92$, $p < .001$), providing strong support for the belief reversal hypothesis (Eq. (6)). We pause again to emphasize what we have found in the Georgia Tech–Washington State conditions: most people (59%) thought that Georgia Tech was *more likely* to win its game than Washington State, whereas most people (76%) assigned a higher *probability* to Washington State winning than Georgia Tech.

DISCUSSION

The preceding surveys provide strong evidence for a familiarity bias in relative likelihood judgment that diminishes or disappears in probability judgment. The familiarity bias violates both probability theory and support theory, which require that hypothesis H is rated more likely than hypothesis L if and only if *not*-L is rated more likely than *not*-H. Results of our studies are summarized in Table 1. Subjects judged a more familiar event to be *more likely* than a less

TABLE 1

Summary of Results

Study	Elicitation mode	$\Pi(H >_i L)$	N	$\Pi(\overline{L} >_i \overline{H})$	N	z
1. Duke sports	More likely	.75	69	.56	66	2.37
2. Mutual funds	More likely	.38	115	.27	113	1.79
3. Corporate theft	More likely	.41	71	.21	73	2.65
4. Temperatures	More likely	.50	84	.37	81	1.70
	Less likely	.50	82	.38	86	1.58
5. Academy Awards	More likely	.76	25	.52	31	1.94
	Less likely	.80	30	.43	30	3.18
6. Law case	More likely	.66	62			
	Probability	.49	55			
		>.22 =.55 <.24				
7. Politics	More likely	.64	22	.24	21	2.89
	Probability	.36	21	.27	22	0.64
		>.29 =.14 <.57		>.18 =.18 <.64		
8. College basketball	More likely	.59	79	.26	76	4.41
	Probability	.24	73	.36	77	−1.62
		>.19 =.10 <.71		>.30 =.12 <.58		

Note. The first column lists the number and topic of the study. The second column lists the elicitation mode. The third column lists the proportion of respondents rating the high familiarity hypothesis (H) above the low familiarity hypothesis (L). The fourth column lists the sample size on which that proportion is based. The fifth column lists the proportion of respondents rating the complement of the high familiarity hypothesis ($\overline{H}$) above the complement of the low familiarity hypothesis ($\overline{L}$). The sixth column lists the sample size on which that proportion is based. The final column lists the z score of the difference between the proportions reported in the third and fifth columns. Extra values in the third and fifth columns are the raw proportions, including ties, from which the reported proportions were derived for judged probabilities.

familiar event more frequently than they judged the complement of the less familiar event to be *more likely* than the complement of the more familiar event. Orderings of beliefs over events were the same when we reversed polarity of the elicitation mode so that subjects were asked which event was *less likely*. Hence, subjects judged a less familiar event *less likely* than a more familiar event more frequently than they judged the complement of the more familiar event to be *less likely* than the complement of the less familiar event. The familiarity bias was greatly reduced when subjects were asked to judge the *probabilities* of these events. We demonstrated these effects in eight studies involving sports, investments, crime, weather, entertainment, legal issues, and politics, with 1464 participants who were undergraduates, business students, law students, members of the Duke University community, and video store patrons. We conclude with a discussion of the implications of these findings for the study of judgment under uncertainty, comments on related work by others, and suggestions for future research.

One of the fundamental assumptions of rational choice theory is *procedure invariance*, according to which normatively equivalent elicitation procedures should produce the same preference ordering. Although researchers have documented robust violations of this principle when establishing preference orderings over multiattribute and risky options (e.g., Tversky, Sattath, & Slovic, 1988), little attempt has been made thus far to study the effects of elicitation mode on belief orderings over events. The present investigation provides compelling evidence that the mode by which beliefs are elicited can affect their ordering in systematic and predictable ways. First, we have presented examples of belief reversals in which a more familiar event was deemed *more likely* by most subjects, but most subjects assigned it a lower *probability*. Second, we have documented examples in which most people judge a familiar event *more likely* to occur than an unfamiliar event *and* most people rate the familiar event *more likely* **not** to occur than the unfamiliar event. This, too, might be interpreted as a belief reversal.

The familiarity bias should be distinguished from research on the *recognition heuristic*, according to which "if one of two objects is recognized and the other is not, then [people] infer that the recognized object has the higher value with respect to the criterion" (Goldstein & Gigerenzer, 1999, p. 7).[8] First, the present account applies to differential familiarity with objects that are, for the most part, recognized; the recognition heuristic, in contrast, is an "all-or-none distinction—degrees of further knowledge are irrelevant" (Goldstein & Gigerenzer, 1999, p. 10). Moreover, the recognition heuristic would only apply to situations in which recognition is assumed to be correlated with relative likelihood. Hence, the recognition heuristic is not relevant to most of the examples presented in this paper because in these cases either: (a) subjects recognize both events (Problem 1: Duke sports); (b) they are provided with information concerning both events (Problem 3: corporate theft); or (c) recognition is not perceived to

[8] This is the formulation for two-alternative choice tasks; the recognition heuristic can also be formulated for more general situations (Goldstein, personal communication).

be diagnostic of relative likelihood (Problem 4: temperatures). Second, the recognition heuristic cannot accommodate situations in which *both* the familiar event and its complement are judged more likely (Problem 2: mutual funds). Finally, the recognition heuristic is mute concerning the observed differences between *more likely* and *probability* elicitation modes because it is not formulated for probability judgment (Goldstein, personal communication).

In order to motivate our predictions concerning the familiarity bias, we have proposed a contingent weighting model in which evidence for the alternative hypothesis looms larger in judgments of probability than in judgments of relative likelihood. The results of our studies are consistent with the predictions generated by such a model. A more direct parametric test of this model can be found in Fox and Levav (2000). Furthermore, in our studies we have manipulated familiarity without attempting to control, in any single survey, for other features of events that might covary with familiarity, such as event importance or self-relevance. The robustness of the reported effects across eight variations, however, provides strong support for the familiarity account.

In this paper we have compared the ordering of beliefs over events when subjects are asked to evaluate which is more or less likely versus estimate the probabilities of each event. We have interpreted the difference in these elicitation modes in terms of the distinction between *relative* and *absolute* likelihood judgment. However, these modes differ in two additional respects. First, assessment of which event is more or less likely requires a *qualitative* judgment, whereas assessment of which event has a higher probability entails *quantitative* judgments. In continuing research (Fox & Levav, 2000) we find that even when subjects are asked to assess the degree of relative likelihood quantitatively on a likert scale, they exhibit a bias in favor of the focal hypothesis that is less pronounced in probability judgment. Second, assessment of which event is more or less likely inherently entails a comparison of a pair of events, whereas probability judgment entails an evaluation of a single event. In Problems 6–8 we have controlled for this factor by asking all subjects in the probability condition to evaluate *both* events in the same context.

Future research might examine the effects of joint versus separate evaluation of probabilities. When judging the probabilities of related events in the same context, new information is introduced by virtue of this juxtaposition. First, the inclusion relationship among events becomes more salient. One might expect greater additivity of probabilities and more extensional reasoning when events are juxtaposed (cf. Tversky & Kahneman, 1983). Although violations of the inclusion rule due to reasoning by representativeness sometimes persist even with "transparent" tests in which participants rank or judge probabilities of target events consecutively (e.g., "Linda is a bank teller" versus "Linda is a bank teller who is active in the feminist movement"), we suspect that more generally such errors will diminish when events are judged together rather than separately.

Second, judging two events in the same context can provide information that aids in the evaluation of diagnostic cues. In studies of multiattribute choice, Hsee (1996) has shown that the preference ordering over options can be reversed when priced jointly versus separately, under conditions in which an important attribute is difficult to evaluate in isolation. For example, one group of subjects priced a music dictionary with 10,000 entries and no defects higher than a second group of subjects priced a music dictionary with 20,000 entries and a torn cover; however, a third group who evaluated the dictionaries jointly priced the latter higher than the former (for a comprehensive review of this literature, see Hsee, Loewenstein, Blount, & Bazerman, 1999). Following Hsee (1996), Fox, Levav, and Payne (2000) asked participants to judge the probability that particular teams would qualify for the National Collegiate Athletic Association (NCAA) basketball tournament based on their record mid-way through the season and their ratings percentage index (RPI), which is a measure used by the NCAA that is weighted heavily by strength of schedules. When separate groups of respondents were asked to make judgments, the mean reported probability was slightly higher for the group evaluating the team with the better record; when a single group of subjects was asked to make judgments concerning both teams, the team with the higher RPI was judged significantly more likely to qualify. It seems that RPI received greater weight when respondents were provided with more information concerning its distribution by virtue of the juxtaposition of teams.

In neoclassical economics, direct expressions of belief are generally regarded with suspicion. Instead, belief orderings over events are established through choices between prospects whose consequences are contingent on these events (e.g., Ramsey, 1931). For instance, when a person prefers to receive $100 if the Denver Broncos win next year's Super Bowl to $100 if the San Francisco 49ers win, we infer that this person believes the former event to be more likely than the latter. Strategically equivalent elicitation modes are assumed to provide the same ordering of subjective probabilities. However, recent psychological research suggests that the major assumptions of the classical theory that underlie the derivation of belief from preference are not descriptively valid and that decisions under uncertainty can be predicted more accurately from direct judgments of probability than from subjective probabilities derived from choices (see, e.g., Fox & Tversky, 1998). Further research might explore the impact of elicitation mode effects in *judgment* as they are manifested in *decision making* under uncertainty. We suspect, for example, that *choices* between two uncertain prospects offering the same outcome contingent on different events (e.g., the football example above) may naturally lead decision makers to make an ordinal evaluation of which event is "more likely," whereas *pricing* uncertain prospects may naturally lead decision makers to form a cardinal evaluation of the likelihood of each event.

The study of choice under uncertainty has shown that people typically prefer to bet on known probabilities over unknown probabilities (Ellsberg, 1961) or, more generally, in areas in which they feel knowledgeable or competent to areas in which they feel ignorant or incompetent (Heath & Tversky, 1991). Moreover, this effect seems to diminish or disappear in the absence of an explicit comparison with other sources of uncertainty that the decision maker feels more or less competent judging or with other people who are more or less knowledgeable (Fox & Tversky, 1995; see also Fox & Weber, 2000). Researchers typically demonstrate competence effects by showing that a person favors a bet on event H to a bet on event L, and also favors a bet on event $\overline{H}$ to a bet on event $\overline{L}$. It is tempting to speculate that the familiarity bias in relative likelihood judgment may contribute to this preference to bet both for and against more familiar events. For instance, when making choices among competing financial investments about which the decision maker has differential familiarity, it may naturally occur to that person to ask himself or herself or an advisor which is "more likely" to be profitable. Such questioning may lead the decision maker to favor an investment in the domestic stock market (cf. Kilka & Weber, 1998) or the telephone company in that person's home state (Huberman, 1998). However, if one were to ask the less natural question of which investment is more likely to yield a *poor* return, the opposite conclusion might be reached.[9]

Regardless of the impact of the familiarity bias in choice under uncertainty, the study of direct expressions of belief is worthy in its own right. We often solicit the opinions of others concerning future events on which our well-being is contingent. We ask doctors, lawyers, financial advisors, and a host of other professionals for their assessments concerning the relative likelihood that various medical treatments will be successful, that various legal ploys will allow us to prevail in court, or that various investments will be profitable. The present findings suggest that the answer we receive may depend crucially on the way in which the question is posed: "which treatment is more likely to succeed?" may yield a different ordering than "which treatment is more likely to fail?" which, in turn, may yield a different ordering than "what is the probability that each treatment succeeds?" The role of elicitation mode in more natural contexts awaits further empirical investigation.

APPENDIX 1: SUFFICIENT CONDITIONS FOR EQS. (4) AND (6)

Sufficient Conditions for Eq. (4).

Assuming:

(1) The likelihood that a person ranks H over L in response mode i is a strictly increasing function of the difference between terms on either side of the inequality in the contingent weighting model (Eq. 3), $[\alpha_i \log s(H) - \beta_i \log s(\overline{H})] - [\alpha_i \log s(L) - \beta_i \log s(\overline{L})]$;

(2) The focal hypothesis receives greater weight than the alternative hypothesis (i.e., $\alpha_\mu > \beta_\mu$);

[9] In connection with this point we speculate that the natural polarity of some events (e.g., "who is more likely to win the tournament?") may predispose people to spontaneously reframe the complementary event (e.g., "who is more likely not to win the tournament?") into a more natural formulation (e.g., "who is less likely to win the tournament?") which would diminish the observed effect sizes of the familiarity bias and belief reversal phenomena.

(3) The sum of log-support for hypothesis H and its complement is larger than the sum of log-support for hypothesis L and its complement, $\log s(H) + \log s(\overline{H}) \geq \log s(L) + \log s(\overline{L})$,

$$\textit{then } \Pi(H >_{\mu} L) \geq \Pi(\overline{L} >_{\mu} \overline{H}).$$

Sufficient Conditions for Eq. (6).

Assuming:

(1) The likelihood that a person ranks H over L in response mode i is a strictly increasing function of the difference between terms on either side of the inequality in the contingent weighting model (Eq. 3),

$$[\alpha_i \log s(H) - \beta_i \log s(\overline{H})] - \alpha_i \log s(L) - \beta_i \log s(\overline{L})];$$

(2) The focal hypothesis receives more weight and the alternative hypothesis receives less weight in "more likely" judgment than in probability judgment: $\alpha_{\mu} > \alpha_P$ and $\beta_{\mu} < \beta_P$;

(3) Support for both the focal and alternative hypotheses is higher for the more familiar event, $s(H) \geq s(L)$, $s(\overline{H}) \geq s(\overline{L})$,

$$\textit{then } \Pi(H >_{\mu} L) \geq \Pi(H >_P L).$$

However, if we assume that the total weight in the contingent weighting model attached to the focal and alternative hypotheses is constant (i.e., $\alpha_i + \beta_i = C$) then (2) becomes more trivial (because $\alpha_{\mu} > \alpha_P$ iff $\beta_{\mu} < \beta_P$) and we can replace (3) with the less restrictive requirement that the sum of log-support for hypothesis H and its complement is larger than the sum of log-support for hypothesis L and its complement, $\log s(H) + \log s(\overline{H}) \geq \log s(L) + \log s(\overline{L})$.

REFERENCES

Brenner, L. A., & Rottenstreich, Y. (1999). Focus, repacking, and the judgment of group hypotheses. *Journal of Behavioral Decision Making*, **12,** 141–148.

Budescu, T., & Wallsten, T. (1995). Processing linguistic probabilities: General principles and empirical evidence. In J. Busemeyer, D. L. Medin, and R. Hastie (Eds.), *Decision making from a cognitive perspective*. New York: Academic Press.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, **75,** 643–669.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, **50,** 123–129.

Fischer, G., Carmon, Z., Ariely, D., & Zauberman, G. (1999). Goal-based construction of preferences: Task goals and the prominence effect. *Management Science*, **45,** 1057–1075.

Fiske, S. E., & Taylor, S. T. (1991). *Social cognition*. New York: McGraw-Hill.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, **46,** 199–210.

Fox, C. R. (1999). Strength of evidence, judged probability, and choice under uncertainty. *Cognitive Psychology*, **38,** 167–189.

Fox, C. R., & Irwin, J. R. (1998). The role of context in the communication of uncertain beliefs. *Basic and Applied Social Psychology*, **20,** 57–70.

Fox, C. R., & Levav, J. (2000). *Contingent weighting and belief reversal in absolute versus relative likelihood judgment*. Unpublished manuscript, Durham, NC: Duke University, Fuqua School of Business.

Fox, C. R., Levav, J., & Payne, J. W. (2000). *Evaluability and belief reversal*. Unpublished manuscript, Durham, NC: Duke University, Fuqua School of Business.

Fox, C. R., & Tversky, A. (1998). A Belief-based approach to decision under uncertainty. *Management Science*, **44,** 879–895.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics*, **110,** 585–603.

Fox, C. R., & Weber, M. (2000). *Ambiguity aversion, comparative ignorance, and context*. Unpublished manuscript, Durham, NC: Duke University, Fuqua School of Business.

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

Goldstein, D. G., & Gigerenzer, G. (1999). *Models of ecological rationality: The recognition heuristic*. Unpublished manuscript, Berlin, Germany: Max Planck Institute for Human Development.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomenon. *Psychological Review*, **94,** 236–254.

Griffin, D., & Tversky, A. (1992). The weighting of evidence and the determinants of confidence. *Cognitive Psychology*, **24,** 411–435.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, **4,** 4–28.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals for joint and separate evaluations of alternatives. *Organization Behavior and Human Decision Processes*, **67,** 247–257.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, **125,** 576–590.

Huberman, G. (1998). *Familiarity breeds investment*. Unpublished manuscript, New York: Columbia University.

Kahneman, D., Slovic, P., & Tversky A. (Eds). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge Univ. Press.

Kahneman, D., Slovic, P., & Tversky A. (Eds). (2000). *Choices, values, and frames*. Cambridge, UK: Cambridge University Press.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan.

Kilka, M., & Weber, M. (1998). *Home bias in international stock return expectations*. Unpublished manuscript, Mannheim, Germany: University of Mannheim.

Macchi, L., Osherson, D., & Krantz, D. H. (1999). A note on superadditive probability judgment. *Psychological Review*, **106,** 210–214.

Mellers, B. A., Ordóñez, L. D., & Birnbaum, M. H. (1992). A change-of-process theory for contextual effects and preference reversals in risky decision making. *Organization Behavior and Human Decision Processes*, **52,** 331–369.

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

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge Univ. Press.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays by F. P. Ramsey*. New York: Harcourt, Brace.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review*, **104,** 406–415.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition*, **21,** 546–556.

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review*, **73,** 596–605.

Tversky, A. (1977). Features of similarity. *Psychological Review*, **84,** 327–352.

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch and B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **91,** 293–315.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, **101,** 547–567.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, **95,** 371–384.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review*, **80,** 204–217.

Wickens, C. D. (1984). *Engineering psychology and human performance*. Columbus, OH: Merrill.

Received July 12, 1999