Psychological Review

VOLUME 92 NUMBER 4 OCTOBER 1985

Ambiguity and Uncertainty in Probabilistic Inference

Hillel J. Einhorn and Robin M. Hogarth
University of Chicago, Graduate School of Business, Center for Decision Research

Ambiguity results from having limited knowledge of the process that generates outcomes. As Ellsberg (1961) demonstrated, this poses problems for theories of probability operationally defined via choices amongst gambles. A descriptive model of how people make judgments under ambiguity is proposed. The model assumes an anchoring-and-adjustment process in which an initial estimate provides the anchor, and adjustments are made for what might be. The latter is modeled as the result of a mental simulation process that reflects two factors: (a) the amount of ambiguity, which affects the size of the simulation, and (b) one's attitude toward ambiguity, which affects the differential weighting of imagined probabilities. A two-parameter model of this process is shown to be consistent with Ellsberg's original paradox, the nonadditivity of complementary probabilities, current psychological theories of risk, and Keynes's idea of the "weight of evidence." The model is tested in four experiments using individual and group analyses. In Experiments 1 and 2, the model accurately predicts inferential judgments; in Experiment 3, the model predicts choices between gambles; and in Experiment 4 it shows how buying and selling prices for insurance are influenced by one's attitude toward ambiguity. The results are then discussed with respect to (a) the importance of ambiguity in assessing uncertainty, (b) the use of cognitive strategies in judgments under ambiguity, (c) the role of ambiguity in risky choice, and (d) extensions of the model.

The literature on how people make judgments under uncertainty is large, complex, and rife with controversy (see, e.g., Cohen, 1981; Edwards, 1954, 1968; Einhorn & Hogarth, 1981; Kahneman, Slovic, & Tversky, 1982; Kyburg, 1983; Peterson & Beach, 1967; Rappoport & Wallsten, 1972; Slovic & Lichtenstein, 1971; Slovic, Lichtenstein, & Fischhoff, 1977). One reason for the controversy is that although there is agreement that uncertainty is a crucial factor in inference, there is much less agreement about its meaning and measurement (cf. Tversky & Kahneman, 1982). In particular, although most psychological work on inference has been guided by a Bayesian or subjectivist view of probability, increasing concerns have been expressed about this position (e.g., Cohen, 1977; Shafer, 1978). Central to the Bayesian view is the idea that probability, which is a measure of one's degree of belief, can be operationally defined via choices amongst gambles (Savage, 1954). Thus, if two gambles have identical payoffs but one is preferred to the other, it follows that the probability of winning is greater for the chosen alternative.

Although the subjectivist view of probability gains much of its force by operationally defining uncertainty via choices amongst gambles,

This research was supported by a contract from the Office of Naval Research.

We wish to thank John Hand, Haim Mano, Ann McGill, and Diane Vitiello for their assistance on this project. In addition, the following people provided helpful comments on earlier versions of the manuscript: Tom Dyckman, Zvi Gilula, William Goldstein, Steven Hoch, Joshua Klayman, Howard Kunreuther, John Payne, Nils-Eric Sahlin, Amos Tversky, Thomas S. Wallsten, and five anonymous reviewers.

Requests for reprints should be sent to Hillel J. Einhorn or to Robin M. Hogarth, Center for Decision Research, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637.

do probabilities derived from choices capture the essential *psychological* aspects of uncertainty? An important and direct attack on this assumption was put forward by Daniel Ellsberg (1961), and we examine his arguments in this article. In doing so, however, we stress that our intent is to understand the psychological bases of uncertainty rather than to critique the normative status of the Bayesian position (cf. Ellsberg, 1963; Raiffa, 1961; Roberts, 1963).

Ellsberg (1961) used the following example to show that the uncertainty people experience contains several aspects, one of which is not captured by probabilities: Imagine two urns, each containing red and black balls. In Urn 1, there are 100 balls, but the proportions of red and black are unknown; Urn 2 contains 50 red and 50 black balls. Now consider a gamble such that, if you bet on red and it is drawn from the urn you get $100; similarly for black. However, if you bet on the wrong color, the payoff is $0. Imagine having to decide which color to bet on if a ball is to be drawn from Urn 1; that is, the choices are red (R_1), black (B_1), or indifference (I). What about the same choices in Urn 2: (R_2), (B_2), or (I)? Most people are indifferent in both cases, suggesting that the subjective probability of red in Urn 1 is the same as the known proportion in Urn 2—namely, .5. However, would you be indifferent to betting on red if Urn 1 were to be used versus betting on red using Urn 2 (R_1 vs. R_2)? Similarly, what about B_1 versus B_2? Many people find that they prefer R_2 over R_1, even though their indifference judgments within both urns imply that $p(R_1) = p(R_2) = .5$. Furthermore, the same person who prefers R_2 over R_1 may also prefer B_2 over B_1. This pattern of response is inconsistent with the idea that even a rank order of probabilities can be inferred from choices. Thus, if R_2 is preferred over R_1, this implies that $p(R_2) > p(R_1)$. Moreover, because red and black are complementary events, this means that $p(B_2) < p(B_1)$. However, if B_2 is preferred over B_1, then $p(B_2) > p(B_1)$, which contradicts the preceding inequality. It is also important to note that if $p(R_2) > p(R_1)$ and $p(B_2) > p(B_1)$, then either Urn 2 has complementary probabilities summing to more than 1 (superadditivity) *or* Urn 1 has complementary probabilities summing to less than 1 (subadditivity). Although Ellsberg did not specifically discuss the nonadditivity of complementary probabilities (cf. Fellner, 1961), we show that it is intimately related to the effects of different types of uncertainty on probabilistic judgments.

From our perspective, the importance of Ellsberg's paradox lies in the difference in the nature of the uncertainty between Urns 1 and 2. In Urn 1, whereas one's best estimate of the probability may be .5, confidence in that estimate is low. In Urn 2, on the other hand, one is at least certain about the uncertainty in the urn. Although it may seem strange, and even awkward, to speak of uncertainty as being more or less certain itself, such a concept captures an important aspect of how people make inferences from unknown, or only partially known, generating processes. Indeed, the idea of uncertainty about uncertainty has been considered from time to time under the rubrics, *second-order* uncertainty and probabilities for probabilities (e.g., Marschak, 1975). However, whereas this concept has received little support amongst subjectivist statisticians (see, e.g., de Finetti, 1977), its status as a psychological factor of importance for understanding choice and inference has been demonstrated experimentally (Becker & Brownson, 1964; Yates & Zukowski, 1976). On the other hand, the process by which such second-order uncertainty is used in inference, and the factors that affect its use, have not been systematically studied. To be sure, Ellsberg (1961) suggested a number of variables that should affect the ambiguity of a situation, including the amount, type, reliability, and degree of conflict in the available information. Indeed, he stated,

> Ambiguity is a subjective variable, but it should be possible to identify 'objectively' some situations likely to present high ambiguity, by noting situations where available information is scanty or obviously unreliable or highly conflicting; or where expressed expectations of different individuals differ widely; or where expressed confidence in estimates tends to be low. Thus, as compared with the effects of familiar production decisions or well-known random processes (like coin-flipping or roulette), the results of Research and Development, or the performance of a new President, or the tactics of an unfamiliar opponent are all likely to appear ambiguous. (pp. 660–661)

To specify the concept of ambiguity more precisely, reconsider the urn where the proportion of red and black balls is unknown. Now consider the set of all possible probability distributions over the proportions of red and black balls. Do we have any reason for thinking that some of these distributions are more likely than others? Without further information, it would seem that the answer is no. In fact, one might consider our lack of knowledge regarding such a situation as representing ignorance, inasmuch as there is no information that allows us to rule out possible distributions (cf. Levi, 1980, on the various ways of representing ignorance in probabilistic systems). However, imagine that a person sampled four balls (without replacement) and got three red and one black. Note that this result rules out certain probability distributions, whereas making others more likely. Thus, the proportion of red is now restricted to $.03 + x$ (where $0 \leq x \leq .96$) and the proportion of black to $.97 - x$. Moreover, as sample size increases, further distributions are ruled out until only one is left. One can now distinguish between ignorance, ambiguity, and risk according to the degree to which one can rule out alternative distributions; that is, ambiguity is an intermediate state between ignorance (no distributions are ruled out) and risk (all but one distribution is ruled out). Thus, ambiguity results from the uncertainty associated with specifying *which* of a set of distributions is appropriate in a given situation. Moreover, the amount of ambiguity is an increasing function of the number of distributions that are not ruled out (or made implausible) by one's knowledge of the situation.

It is important to note that sample size is only one factor that influences ambiguity because other information can affect one's assessment of which probability distributions should be eliminated from consideration. Thus, imagine an urn factory where employees color balls by throwing them at two adjacent cans of black and red paint from a distance of 20 feet. Given our knowledge of this process, it seems fair to expect that an urn of 100 balls would not contain extreme proportions of red or black. A second example, due to Gärdenfors and Sahlin (1982), is particularly illuminating on this issue:

> Consider Miss Julie who is invited to bet on the outcome of three different tennis matches. As regards match A, she is very well-informed about the two players. . . . Miss Julie predicts that it will be a very even match and a mere chance will determine the winner. In match B, she knows nothing whatsoever about the relative strength of the contestants. . . and has no other information that is relevant for predicting the winner of the match. Match C is similar to match B except that Miss Julie has happened to hear that one of the contestants is an excellent tennis player, although she does not know anything about which player it is, and that the second player is indeed an amateur so that everybody considers the outcome of the match a foregone conclusion. (pp. 361–362)

Note that the amount and type of information in the three situations is quite different, as is the amount of ambiguity. We would argue that Match A is the least ambiguous because the number of reasonable distributions is small (e.g., the probability of a given player winning might be as low as .49 and as high as .51). Match B, on the other hand, could involve all possible distributions over the probability of winning (analogous to the urn with the unknown proportions of colored balls). Match C is an interesting case that we believe is closer in ambiguity to A than to B. The reason is that the distributions for a given player winning the match are limited; that is, the probability of winning might be either $1 - x$ (where $0 \leq x \leq .05$) or x. From our perspective, how does the amount and type of ambiguity affect judgments of the probability of winning or losing the match? Would Miss Julie, for example, judge that each player in the three matches has a .5 chance of winning (or losing)?

Our discussion so far has implied that ambiguity is generally avoided because it adds to the total uncertainty of a situation. Indeed, this is explicitly mentioned by Ellsberg (1961, p. 666) in discussing why new technologies are resisted more than one would expect on the basis of their first-order probabilities. However, this picture is not completely accurate, as is made clear by another Ellsberg example (as quoted in Becker & Brownson, 1964, pp. 63–64, Footnote 4): Consider two urns with 1,000 balls each. In Urn 1, each ball is numbered from 1 to 1,000, and the probability of drawing any number is .001. In Urn 2, there are an unknown number of balls bearing any single number. Thus, there may be 1,000 balls with number 687, no balls with this number, or anything in between. If there is a prize for drawing Number 687 from the urn, would you prefer to draw from Urn 1 or Urn 2? Note that Urn 1 has no ambiguity and each numbered ball has the same .001 chance of being drawn. Urn 2, on the other hand, can be characterized as inducing extreme ambiguity (i.e., ignorance). However, for many people, the drawing from Urn 2 seems considerably more attractive than from Urn 1, thereby implying that there are situations in which ambiguity is preferred rather than avoided. This is considered in detail later in this article, but we note here that accounting for such shifts is an important criterion for judging the adequacy of any theory of inference under ambiguity.

Finally, the concepts of ambiguity, second-order uncertainty, and the like have been of concern in theories of inference, quite apart from their role in affecting choice. For example, work on fuzzy sets (Zadeh, 1978), Shafer's (1976) theory of evidence, Cohen's (1977) attempt to formalize uncertainty in legal settings, and the elicitation of probability ranges (Wallsten, Forsyth, & Budescu, 1983) all contain ideas concerning the vagueness that can underlie probabilities. Indeed, statisticians have provided axiomatic systems for trying to formalize probability ranges and rank orders rather than specific values (e.g., Koopman, 1940). Moreover, early work by Keynes (1921) addressed the notion of ambiguity by distinguishing between probability and what he called the "weight of evidence." Keynes (1921) stated,

> The magnitude of the probability . . . depends upon a balance between what may be termed the favourable and the unfavourable evidence; a new piece of evidence which leaves this balance unchanged, also leaves the probability of the argument unchanged. But it seems that there may be another respect in which some kind of quantitative comparison between arguments is possible. This comparison turns upon a balance, not between the favourable and unfavourable evidence, but between the *absolute* amounts of relevant knowledge and of relevant ignorance respectively. (p. 71)

A Descriptive Model

We first present a *descriptive* model of how people make probability judgments and choices under varying amounts of ambiguity. We require that our model be able to (a) explain the pattern of choices elicited by Ellsberg's problems. This, in turn, implies that the model account for sub- *and* superadditivity of the probabilities of complementary events; (b) specify the conditions under which people seek as well as avoid ambiguity; (c) allow for individual differences; and (d) be empirically testable. To meet these criteria, the model is tested in four experiments at both the aggregate and individual subject levels. Two of the experiments concern inference tasks, and two involve choices. The implications of the theory and empirical work are then discussed in relation to (a) the importance of ambiguity in assessing perceived uncertainty, (b) the use of cognitive strategies in understanding probabilistic judgments under ambiguity, (c) the role of ambiguity in risky choice, and (d) extensions of the model to multiple sources and time periods.

Anchoring-and-Adjustment Strategy

Our model postulates an *anchoring-and-adjustment* strategy for assessing probabilities. This involves an initial assessment denoted p_A and an adjustment to reflect the ambiguity in the situation. The initial estimate may be obtained from a variety of sources; for example, it could be the best guess of experts, reflect one's previous information about the topic, or it could simply be a number that is salient in memory. For our purposes, the initial estimate permits a starting point or anchor from which adjustments are made. We denote the judgment that results from the anchoring-and-adjustment strategy as $S(p_A)$; thus,

$$S(p_A) = p_A + k, \qquad (1)$$

where k is the net effect of the adjustment process. To model the adjustment process, we propose that people engage in a mental simulation in which other values of p are considered by imagining how well they express one's uncertainty (cf. Kahneman & Tversky, 1982). The rationale for the simulation is that in ambiguous situations, p can come from any one of a number of distributions. Thus, simulating various values of p can allow one to assess which distributions are more or less plausible. These simulated values are then incorporated into the adjustment term, thereby allowing people to maintain sensitivity to both uncertainty and ambiguity.

We further argue that k, the net effect of the simulation, is affected by three factors:

1. The level of p_A; that is, because $S(p_A)$ varies between 0 and 1, Equation 1 implies that $-p_A \le k \le (1 - p_A)$. This means that the direction of the adjustment must be due, in part, to the value of p_A. Indeed, when $p_A = 0$, $k \ge 0$, and the adjustment (if there is one) must be upward; when $p_A = 1$, $k \le 0$, so that the

adjustment must be downward; when $p_A \neq 0$, 1, adjustments can be up or down.

2. The amount of ambiguity perceived in the situation. This affects the absolute size of the adjustment that is captured by a parameter θ ($0 \leq \theta \leq 1$); that is, the greater the perceived ambiguity, the larger the adjustment.

3. The person's attitude toward ambiguity in the circumstances. This is reflected in the tendency to give differential attention or weight to values of p that are greater or smaller than the initial estimate, p_A. Attitude toward ambiguity is denoted by β, and this parameter, together with p_A, determines the sign of the net effect of the adjustment (i.e., when k is positive or negative).

To model the adjustment process algebraically, let

$$k = k_g - k_s, \tag{2}$$

where k_g denotes the effect of imagining values of p greater than the initial estimate, and k_s the effect of imagining smaller values. How does perceived ambiguity affect these quantities? To answer this, consider Figure 1, which shows the position of p_A relative to the end points of 0 and 1. Note that the maximum upward adjustment to p_A is $(1 - p_A)$ and the maximum downward adjustment is p_A. Thus, maximum adjustments would occur when $\theta = 1$ and no adjustments when $\theta = 0$. This suggests that the effects of simulating values greater and smaller than p_A (k_g and k_s, respectively), can be represented as proportions of the maximum adjustments where θ is the constant of proportionality; that is,

$$k = \theta(1 - p_A) \tag{3a}$$

and

$$k_s = \theta p_A. \tag{3b}$$

The development so far ignores the possibility that greater and smaller values than p_A could be differentially weighted. For example, in estimating the chance of an accident in a new technology (high ambiguity), one may start with the estimate offered by the engineering department and then weight larger values of p(accident) more than smaller ones. To account for differential weighting effects, we need only weight either k_g or k_s to affect k. For convenience, we weight k_s (rather than k_g) by β as follows:

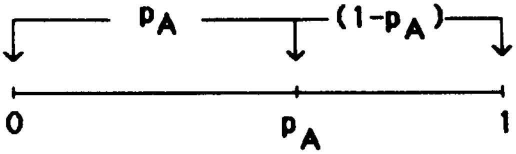

Figure 1. Anchor of p_A and range of adjustments.

$$k_s = \theta p_A^{\beta} \ (\beta \geq 0). \tag{4}$$

Thus, the net effect of the adjustment process is given by

$$k = k_g - k_s$$
$$= \theta(1 - p_A - p_A^{\beta}). \tag{5}$$

When Equation 5 is substituted into Equation 1, the full model becomes

$$S(p_A) = p_A + \theta(1 - p_A - p_A^{\beta}). \tag{6a}$$

Note that the full model can also be expressed as

$$S(p_A) = (1 - \theta)p_A + \theta(1 - p_A^{\beta}). \tag{6b}$$

This form implies that judged probability, $S(p_A)$, is a weighted average of p_A and $(1 - p_A^{\beta})$, where the weights reflect the amount of ambiguity perceived in the situation. Although we prefer the form shown in Equation 6a, we note that Equation 6b is consistent with Lopes's (1981) idea that anchoring-and-adjustment processes often result in outcomes that can be modeled as weighted averages.

We make several points with respect to Equation 6a. First, note that θ affects the absolute size of the adjustment factor. That is, when there is no ambiguity, $\theta = 0$, and $S(p_A) = p_A$. Thus, θ can be thought of as having a magnifying or dampening effect on one's attitude toward ambiguity in the circumstances, (β). For example, if perceived ambiguity is small, the tendency to weight differentially values of p above and below p_A is of little consequence.

Second, $S(p)$ is regressive with respect to p. This can be illustrated by considering the effects of different values of β in Equation 6a. Specifically, the three panels of Figure 2 illustrate *ambiguity functions,* with $\beta < 1$, $\beta > 1$, and $\beta = 1$. (θ is shown to be the same in all three cases.) It is important to note that each value of β defines a unique crossover point, p_c, in which $S(p) = p$. Thus, in Figure 2a, β defines p_{c1} such that small probabilities are overweighted and larger probabilities under-

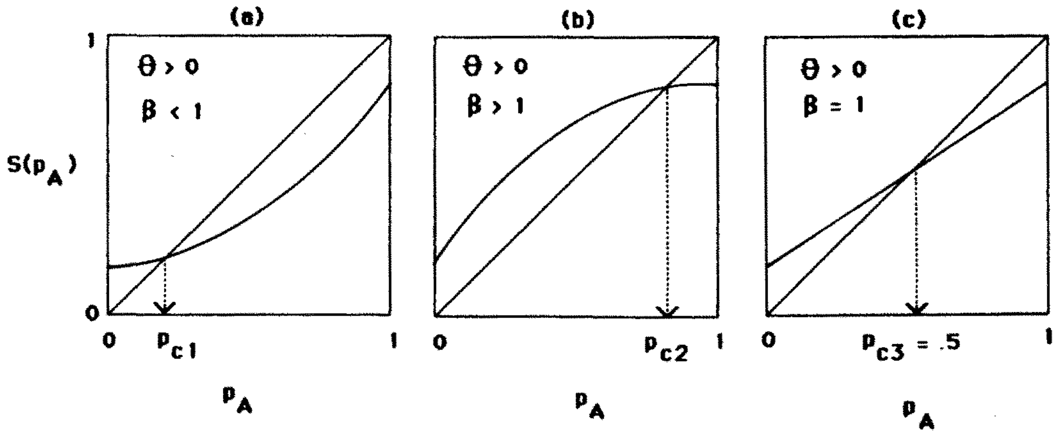

Figure 2. $S(p_A)$ as a function of p_A for values of θ and β.

weighted. This form of the function results because $\beta < 1$ implies that more weight is given to smaller values of p rather than larger ones. Therefore, $k < 0$ over most of the range of p. However, when $p_A < p_{c1}$, there are few smaller values of p to consider relative to larger ones. Thus, even when smaller values are weighted more heavily than larger ones, there are more of the latter, and $k > 0$. Conversely, when $\beta > 1$, as shown in Figure 2b, $S(p) > p$ over most of the range of p because more weight is given to larger as opposed to smaller ps. However, when $p_A > p_{c2}$, $S(p) < p$ because there are few larger values, and $k < 0$. Finally, note that in Figure 2c, when $\beta = 1$, the crossover point is at .5.[1]

Third, Equation 6a implies that the conditions under which probability judgments of complementary events are additive (sum to one). Specifically, consider the sum of $S(p_A)$ and $S(1 - p_A)$. This is,

$$\begin{aligned} S(p_A) &+ S(1 - p_A) \\ &= p_A + \theta(1 - p_A - p_A{}^{\beta}) + (1 - p_A) \\ &\quad + \theta[1 - (1 - p_A) - (1 - p_A)^{\beta}] \\ &= 1 + \theta[1 - p_A{}^{\beta} - (1 - p_A)^{\beta}]. \end{aligned} \tag{7}$$

Thus, complementary probabilities are additive if $\theta = 0$, or $\beta = 1$, or $p_A = 0, 1$; otherwise, there is subadditivity if $\beta < 1$, and superadditivity if $\beta > 1$.

Fourth, there are many ways we could have chosen to incorporate the β parameter in the model. However, not all forms have the same implications, particularly with respect to the additivity of complementary probabilities. We consider several alternative models in the Appendix.

To summarize, the model has two parameters and both are functions of individual and situational factors. The θ parameter reflects perceived ambiguity and the degree to which one simulates values of p that might be. However, situational factors are also likely to affect θ, for example, the absolute amount of evidence available, the unreliability of sources, lack of causal knowledge regarding the process generating outcomes, and so on. The β parameter reflects the extent to which one differentially weights in imagination possible values of p that are smaller versus larger than p_A. As such, β may be related to an optimism–pessimism attitude at the individual level. However, we argue that β is also influenced by situational variables such as the sign and size of the payoffs that are contingent on the ambiguous probability. For example, if the general effect of ambiguity is to induce caution rather than riskiness, the prospect of an undesirable outcome (e.g., monetary losses) would induce people to pay more attention in imagination to values of p(loss) that are larger than p_A; similarly, the prospect of a gain would focus

[1] It is assumed in the above figures that the parameters θ and β are such that $S(p_A)$ is monotonic with p_A. Although Equation 6a is general enough to yield nonmonotonic functions, the empirical data presented later in the article indicate few violations of monotonicity.

attention on smaller values of p(gain). We consider this issue further in connection with Experiments 3 and 4.

We now consider how the model in Equation 6a explains Ellsberg's original results. Note Figure 2a, where $\theta > 0$ and $\beta < 1$. A person with parameter values in these ranges will *underweight* all p_A above p_c, and *overweight* $p_A < p_c$. This particular pattern explains why most people in Ellsberg's urn example avoid the ambiguous Urn 1; that is, $S(p_A = .5) < .50$. However, note that if p_A is less than p_c, $S(p_A) > p_A$, and one would expect the same person who avoided the ambiguous urn when $p_A = .5$, to *prefer* the ambiguous urn when p_A is sufficiently low (e.g., when $p_A = .001$). The pattern of overweighting small p_A and underweighting moderate-to-large p_A also accounts for some otherwise puzzling results of Goldsmith and Sahlin (as reported in Gärdenfors & Sahlin, 1982). They presented subjects with descriptions of either well-known events (e.g., drawing cards from a standard deck), or events about which the subjects had little knowledge (e.g., the likelihood of a bus strike in Verona, Italy, next week). Subjects estimated the probabilities of the events and the perceived reliability of their probability estimates. Events with equal probabilities but unequal reliabilities were then used in a lottery setup. The authors report that "*for probabilities other than fairly low ones* [italics added], lottery tickets involving more reliable probability estimates tend to be preferred" (Gärdenfors & Sahlin, 1982, p. 363).

Although the pattern shown in Figure 2a accounts for much data, it does not explain why some people in the Ellsberg task prefer to bet on drawing from the ambiguous urn when $p_A = .5$. However, consider a person with an $S(p_A)$ function as shown in Figure 2b. When $\theta > 0$ and $\beta > 1$, one gets *ambiguity preference* over most of the range of p_A. Thus, when $p_A < p_c$, $S(p_A) > p_A$ and overweighting occurs; when $p_A > p_c$, $S(p_A) < p_A$ and underweighting occurs. Because individual differences are rarely accounted for in research on decisions under uncertainty, our model has the distinct advantage of positing a general psychological process and allowing for individual differences via particular parameter values. Indeed, this is nicely illustrated by considering people who are indifferent between gambles from ambiguous and unambiguous urns when $p_A = .5$ (as in the Ellsberg case). Our model suggests two distinct types: those for whom $\theta = 0$, and thus, $S(p_A) = p_A$, and those for whom $\theta > 0$ and $\beta = 1$ (shown in Figure 2c). This latter group does not adjust at $p_A = .5$, but does adjust at all other values. Therefore, people characterized by these parameter values will only be indifferent between lotteries at .5.

Finally, we note that our model bears an interesting similarity to the major psychological theory of risk, namely, *prospect theory* (Kahneman & Tversky, 1979). In prospect theory, the effects of uncertainty on choice are modeled via a decision weight function, $\pi(p)$, that is remarkably similar to the $S(p_A)$ function shown in Figure 2a. Although the decision weight function of prospect theory deals with nonambiguous probabilities, Kahneman and Tversky specifically point out that decision weights can be affected by ambiguity. Indeed, they state,

> The decision weight associated with an event will depend primarily on the perceived likelihood of that event, which could be subject to major biases. In addition, decision weights may be affected by other considerations, such as ambiguity or vagueness. Indeed, the work of Ellsberg and Fellner implies that vagueness reduces decision weights. (p. 289)

Thus, although the domain of our theory is different from that of prospect theory, we believe that it is not coincidental that the treatment of uncertainty is so similar. Clearly, further work is necessary to link the two approaches, and current work is proceeding in that direction (Hogarth & Einhorn, 1985).

Experimental Tests of the Model

To test our model empirically, we employed two tasks that focus on inference (Experiments 1 and 2) and two dealing with choice (Experiments 3 and 4). In the inference task, people are asked to make probability judgments on the basis of numbers of reports from a source. In Experiment 1, we examined the various implications of Equation 6a. In Experiment 2, we used different scenarios to manipulate θ in both a between- and within-subjects design. In addition, the consistency of individual differences in strategy (as measured by a person's θ and β parameters) is also considered. Experiment 3 involves an attempt to answer the

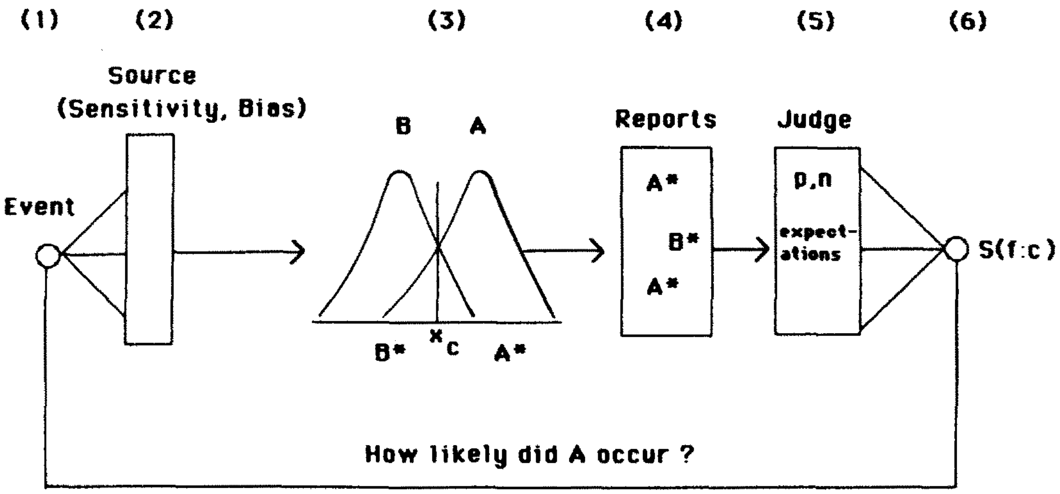

Figure 3. Structure of the experimental inference task.

question: Can an individual's choices between gambles be predicted from knowledge of his or her θ and β parameters obtained from a separate inference task? Finally, in Experiment 4, people are asked to be either buyers or sellers of insurance in ambiguous and nonambiguous situations. Differences between buying and selling prices are then investigated as a function of assumed differences in β parameters. Because Experiments 1–3 are all based on the same type of inference task, we first explicate the underlying nature of this task, noting how it differs from other probabilistic tasks considered in the literature.

Model for Studying Ambiguity in Inference

The prototypical inference that we consider involves a judge assessing the likelihood of the occurrence of an event based on reports received from a source of limited reliability. The task can be thought of as having the elements schematically represented in Figure 3:

1. An event occurs.

2. The event is sensed by observers (e.g., witnesses to an accident) who, in principle, can be characterized by levels of sensitivity and bias. However, it is important to emphasize that these levels are unknown to the judge (see point 5).

3. The observers report what they saw. We denote A* as the report of Event A, and B* as the report of Event B, where the decision rule is to report A* if the observation is above some critical value X_c, and B*, otherwise. The reports can therefore be conceptualized as coming from a signal-detection task.

4. Because there are n observers, n reports are collected. These reports can be thought of as the outcomes of n observers reporting on a single trial of a signal-detection task. Furthermore, because we do not differentiate between the n observers, we refer to them as coming from a single source.

5. The judge receives the information in the form of f reports for a hypothesis (i.e., f reports of A*) and c reports of an alternative (i.e., c reports of B*), where $f + c = n$, and $p = f/n$. The content of the scenario, however, is assumed to give the judge some information as to what values of p to expect in a sample of size n. Specifically, we argue that expectations concerning p are influenced by (a) the dissimilarity between Events A and B and (b) the credibility of the source. By *credibility* we mean the sensitivity and response bias of the observers in judging the particular events of interest. For example, imagine that you are a detective investigating a bank robbery where two witnesses claim that the robber has blond hair and one witness claims it is brown. How likely is it that the robber does have blond hair? Although the detective knows neither the hit and false-alarm rates of the witnesses nor their response bias for saying "blond" versus

"brown," he or she may know something about the quality of eyewitnesses in a robbery, the confusability of blond and brown in the circumstances, and perhaps something about the motivation of the witnesses. Now contrast a situation where the source is two color-television cameras that were filming the robbery at the bank. Whereas in the former case the detective would expect the reports to conflict (i.e., $0 < p < 1$), in the latter it would be surprising if p were not equal to either 0 or 1.

It is important to note that the judge's assessment of the likelihood of A depends both on the reports observed (i.e., $p = f/n$) and his or her knowledge of the situation. The latter, it should be recalled, can be represented by the possible distributions over A that the judge has not been able to eliminate from consideration and that affect the mental simulation process. Thus, in a highly ambiguous situation, the information about the credibility of the source, the dissimilarity of the signals, and the size of the sample does not rule out many distributions. This, in turn, implies a wide range of p values that the judge simulates in assessing the probability of A once he or she has learned the particular value of $p(= f/n)$ reported. Contrast this with a situation where a highly credible source discriminates between dissimilar signals (e.g., evidence from cameras filming the robbery). In this case, ambiguity would be low because knowledge of the process rules out many distributions over the Event A such that relatively few values of p would be considered in the mental simulation. An interesting situation results when the credibility of the source is particularly low and/or the signals are very similar. For example, imagine a blindfold taste test between Pepsi and Coke for randomly chosen shoppers. If we believe that the two drinks are very similar in taste and that most shoppers cannot tell the difference, we would assess the probability of a randomly chosen shopper correctly identifying Pepsi as being close to .5. However, note that this implies a situation of low ambiguity similar to Match A in the example of Miss Julie given earlier in the article. Thus, whereas some authors have equated increased reliability of evidence with less ambiguity (as suggested by Ellsberg), this example illustrates the point that *decreased* reliability can also lead to low ambiguity. Another way to express this is to note that whereas high reliability implies low ambiguity, low ambiguity does not necessarily imply high reliability.

6. The assessment of the likelihood of A results from the combining of p and n with one's expectations concerning the range of possible values of p.

The structure of this task is both similar to and different from several probabilistic models of the inference process. First, it is similar to cascaded inference in that the judge is making inferences about an event on the basis of unreliable reports (cf. Schum, 1980; Schum & Kelley, 1973). Moreover, in our case the judge does not know the precise value of the source's reliability; rather, there is ambiguity concerning what this is (see also Schum & Martin, 1982).

Second, because each observer can be thought of as participating in the same signal-detection task, the reports not only reflect their sensitivity to competing signals, but also their bias due to differential payoffs. However, as recently emphasized by Birnbaum (1983), the manner in which the judge treats the observer reports depends on some theory about the observers. For example, the observer reports could be responsive to the prior probabilities of A and B, as well as to differential payoffs. We emphasize that in our task the judge is not given precise information about these matters. Furthermore, because the judge only receives information on a single trial, the observers' hit and false-alarm rates are not known. Instead, the observed p, and the judge's expectations about p, become cues to the likelihood that the event occurred.

Third, one might consider our situation as a conventional Bayesian revision task (cf. Edwards, 1968). However, the explicit probabilities necessary to assess the likelihood functions are not provided, and no base-rate data or prior probabilities are stated. It would, of course, be possible to provide the judge with explicit prior probabilities. This would, however, be extending our paradigm to one where multiple sources of information need to be combined (i.e., base rates and individuating information). For the sake of simplicity, we only consider the effects of ambiguity on inferences from a single source and thus do not discuss the effects of explicit base rates. (Extensions of our model to multiple sources is considered in the Discussion section.) On the other hand, there is a sense in which prior probabilities are used in our task. To see this, note that in the Bayesian model, one starts with a prior distribution that is then updated by the data (via the likelihood function). In our model, people may be thought of as entertaining a class of prior distributions based on their knowledge of the situation being considered. However, it is only after they receive data and engage in the mental simulation process that they can rule out (or make implausible) various distributions. Those that remain are incorporated into the overall judgment together with the data. Therefore, the process is one in which data constrain the class of priors, which is then integrated with the data. Although this process is normatively inappropriate (inasmuch as the data affect one's priors), it must be recognized that, under ambiguity, such a process provides a simple way of handling a complex inferential task within information processing limitations.

The model shown in Figure 3 explicitly recognizes three sources of ambiguity, viz: (a) the dissimilarity between Events A and B, (b) the credibility of the source, and (c) the number of reports, or sample size, n. Specifically, when n is small, one would expect ambiguity to be high because few distributions are ruled out; however as n increases, we would expect ambiguity to decrease. Thus, to incorporate the effects of n explicitly in our model, let $\theta = \theta'/n$ such that

$$S(f{:}c) = p_A + \frac{\theta'}{n}(1 - p_A - p_A{}^{\beta}), \quad (8)$$

where $S(f{:}c)$ = judged probability, and $p_A = f/n$. That is, in judging the probability of an event based on f reports *for* and c *con*, people are assumed to anchor on f/n, and then adjust for the unreliability of the source and the amount of data. The model in Equation 8 has several implications:

1. Consider the effect of the amount of information (n) on judged likelihood. Note that $S \rightarrow p_A$ as $n \rightarrow \infty$. This means that as the amount of information increases, one becomes more certain as to the diagnosticity of the data. It is important to realize that as $n \rightarrow \infty$, S does not go to 0 or 1 as would be implied by a standard Bayesian revision model. Instead, the fact that S asymptotes at p_A parallels an analogous result in cascaded inference where, under certain symmetry assumptions and assuming conditional independence, the maximum probability of a hypothesis is bounded by the reliability of the reporting source (Schum & DuCharme, 1971).

2. Conditional on a given value of θ', the model implies that there will be trade-offs between p and n in determining judged likelihood. For example, one might find the evidence in favor of some hypothesis to be more convincing on the basis of 9:1 than 2:0. However, because S asymptotes at p_A, trade-offs of p and n will only occur at small values of n.

3. Because $\theta = \theta'/n$, n also affects the conditions underlying the additivity of complementary probabilities. Specifically,

$$S(f{:}c) + S(c{:}f) = 1 + \frac{\theta'}{n}[1 - p_A{}^{\beta} - (1 - p_A)^{\beta}], \quad (9)$$

where $S(c{:}f)$ is the judged likelihood of the *alternative* hypothesis based on the same data (we adopt the convention that the order of f and c conveys which hypothesis is being judged). Thus, in addition to the additivity conditions discussed in regard to Equation 7, as $n \rightarrow \infty$, additivity will hold regardless of θ', β, or p_A. Of course, when n is small (meager data), adjustments are substantial and violations of additivity are most likely.

Experiment 1 explicitly considers the role of n in Equation 8, whereas factors affecting θ' are the central concern of Experiment 2.

Experiment 1

Method

Subjects. Thirty-two subjects were recruited through an ad in the university newspaper that offered $5 an hour for participation in an experiment on judgment. The median age of the subjects was 24 years, their educational level was high (mean of 4.4 years of formal post-high-school education), and there were 16 men and 16 women.

Stimuli. The stimuli consisted of a set of scenarios that involved a hit-and-run accident seen by varying numbers of witnesses. Moreover, of the n witnesses to the accident, f claimed that it was a green car, whereas c claimed it was a blue car. A typical scenario was phrased as follows:

> An automobile accident occurred at a street corner in downtown Chicago. The car that caused the accident did not stop but sped away from the scene. Of the n witnesses to the accident, f reported that the color of the

offending car was green, whereas c reported it was blue. On the basis of this evidence, how likely is it that the car was green?

Each scenario was printed on a separate page and contained a 0–100-point rating scale that was used by the subject to judge how likely it was that the accident had been caused by a particular color car. Each stimulus contained the same basic story but varied in the total number of witnesses (n), the number saying it was a green (f) or a blue car (c), and whether one was to judge the likelihood that the majority or minority position was true. In order to sample a wide range of values of n and p, 40 combinations were chosen, as follows: for $p = 1$, $n = 2, 6, 12, 20$; $p = .89$, $n = 9, 18, 27$; $p = .80$, $n = 5, 10, 15, 20, 25$; $p = .75$, $n = 4$; $p = .67$, $n = 3, 6, 9, 12, 15, 18, 24$; $p = .60$, $n = 5, 10$; $p = .50$, $n = 2, 8, 12, 20$; $p = .40$, $n = 5, 10$; $p = .33$, $n = 6, 9, 18$; $p = .25$, $n = 4$; $p = .20$, $n = 5, 10$; $p = .11$, $n = 9, 18$; $p = 0$, $n = 2, 6, 12, 20$. In addition, eight stimuli were given twice to ascertain test–retest reliability. Thus, the total number of stimuli was 48, and they were arranged in booklet form.

It is important to note that of the 40 different stimuli (not counting repeats), 10 involved asking the subject for the probability of the alternative hypothesis (i.e., the probability that the car was blue) based on exactly the same data used to assess the probability that the car was green. The responses to these stimuli therefore provide a direct test of the additivity of complementary probabilities. The stimuli for this test involved $p = .89$, $n = 9, 18$; $p = .80$, $n = 5, 10$; $p = .75$, $n = 4$, $p = .67$, $n = 6, 18$; and, $p = .60$, $n = 5, 10$.[2]

Procedure. When the subjects entered the laboratory, they were told that the experiment involved making inferential judgments. Furthermore, it was stated that if they did well in the experiment (without specifying what this meant), it was likely that they would be called for further experiments. Given the relatively high hourly wage, this was thought to provide some incentive to take the task seriously. In order to avoid boredom and to reduce the transparency that judgments of complementary events were sometimes required, subjects were given four sets of 12 stimuli and, after completing each set, they performed a different task. All stimuli were randomly ordered within the four sets. Subjects could take as much time as they needed, and they were free to make as many (or as few) calculations as they wished. After completing the task, all of the subjects filled out a questionnaire regarding various demographic variables.

Estimating the model. To estimate the model from the experimental data, we need to rewrite Equation 8 and include a random error term to represent judgmental inconsistency; therefore,

$$S(f{:}c) = p_A + \frac{\theta'}{n}(1 - p_A - p_A{}^{\beta}) + \epsilon. \qquad (10)$$

Equation 10 requires a nonlinear estimation technique. We used a nonlinear regression program (Dixon, 1983) and chose the criterion of minimizing mean absolute deviations (MAD) for estimating the parameters.[3] This program uses a maximum likelihood procedure and yields unique parameter estimates for θ' and β, together with asymptotically efficient standard errors of these estimates. These statistics permit one to test the hypothesis that the adjustment term in the model is significantly greater than zero.

Results

Before discussing the major results, recall that for each subject, eight stimuli were given twice so that test–retest reliability could be assessed. This was done in two ways: (a) The correlation between judgments of the same stimuli, within each subject ($N = 8$), was computed. The mean of these correlations was .93, with 26 of the 32 coefficients greater than .90. (b) Each subject was considered a replication with eight responses, and the correlation between judgments for identical stimuli, over subjects ($N = 256 = 32$ subjects $\times$ 8 responses) was .91. Clearly, the reliability of the judgments was high, regardless of the computational method.

For a general impression of how well the model fits the data, we first consider an aggregate analysis (individual differences are considered in detail later in this article). For each of the 48 stimuli, the judgments from the 32 subjects were averaged to form the mean judged likelihood, $\bar{S}(f{:}c)$. This was then used as the dependent variable to be fit by the model. The parameter values obtained from the estimation program were $\hat{\theta}' = .38$, ($t = 14.89$, $p < .0001$), $\hat{\beta} = .12$ (implying $\hat{p}_c = .18$; $t = 3.20$, $p < .001$). In addition, the mean absolute deviation of model and data was .021.

To see whether the implications of the model hold, consider Table 1, which shows means and predictions; that is, $\bar{S}(f{:}c)$ and $\hat{S}(f{:}c)$, for the 48 stimuli. First, does $S(f{:}c) \rightarrow p_A$, as n increases? The data strongly support this when $p_A = 1, .67, .60, .50, .40$, and 0. At the values of .75 and .25, n was not varied, although the

[2] In order to more fully test the additivity conditions, we ran a second group of subjects ($N = 24$), where 18 of the 48 stimuli allowed for a direct test. The test stimuli included the 10 listed above, as well as $p = 1$, $n = 2, 6, 12, 20$, and, $p = .5$, $n = 2, 8, 12, 20$. The results for this group were virtually identical to those discussed above.

[3] In addition to using an absolute-deviation–loss function, we also investigated a squared-error–loss function. The results were similar. We also note that the regression program cannot handle values of $p = 0$ when $\beta = 0$ because this means computing 0^0. Therefore, for the purpose of estimation, $p = 0$ was replaced by $p = 10^{-6}$.

Table 1
Fit of the Model for Aggregate Data

n	p_A	$\bar{S}$	$\hat{S}$
2	1	.85	.81
6	1	.92	.94
12	1	.96	.97
20	1	.95	.98
9	.89	.88	.85
18	.89	.87	.87
(18)	(.89)	(.85)	(.87)
27	.89	.87	.88
5	.80	.80	.74
10	.80	.73	.77
(10)	(.80)	(.79)	(.77)
15	.80	.81	.78
20	.80	.80	.79
25	.80	.82	.79
(25)	(.80)	(.80)	.79
4	.75	.63	.68
3	.67	.61	.59
(3)	(.67)	(.59)	(.59)
6	.67	.62	.63
(6)	(.67)	(.63)	(.63)
9	.67	.61	.64
12	.67	.64	.65
15	.67	.65	.65
18	.67	.63	.66
24	.67	.66	.66
5	.60	.53	.56
10	.60	.58	.58
2	.50	.45	.42
8	.50	.44	.48
(8)	(.50)	(.47)	(.48)
12	.50	.47	.49
20	.50	.47	.49
5	.40	.36	.38
10	.40	.39	.39
6	.33	.31	.32
(6)	(.33)	(.29)	(.32)
9	.33	.27	.32
18	.33	.29	.33
4	.25	.20	.24
5	.20	.21	.20
10	.20	.19	.20
(10)	(.20)	(.18)	(.20)
9	.11	.12	.12
18	.11	.13	.11
2	0	.16	.19
6	0	.07	.06
12	0	.06	.03
20	0	.04	.02

Note. Numbers in parentheses are for the repeat judgments. p_A = initial assessment; $\bar{S}$ = mean judgment; $\hat{S}$ = predicted mean judgment.

large adjustments do suggest that the expected effect would occur. However, the effect of n is less clear at p = .89, .80, and .33 because there is little initial adjustment at small n. Taken together, these results suggest moderate support for the hypothesis. Second, do p and n trade off in affecting judged likelihood? The evidence here is quite convincing; for example, note that S(8:1) = .88 > S(2:0) = .85, S(10:5) = .65 > S(3:1) = .63, S(1:4) = .21 > S(1:3) = .20. Of particular interest is the result that S(0:2) = .16 > S(1:8) = .12. This means that when there is limited evidence, no data in favor of a hypothesis can be judged as stronger evidence *for* that hypothesis than when more evidence is available with mixed support. Third, an important implication of the model concerns the relation between θ, β, and the additivity of complementary probabilities. Recall from Equation 7 that when $\theta > 0$ and $\beta < 1$, subadditivity is predicted for $0 < p_A < 1$. To test this prediction, consider Table 2, which shows both $\bar{S}(f{:}c) + \bar{S}(c{:}f)$ and $\hat{S}(f{:}c) + \hat{S}(c{:}f)$. Note that there is substantial subadditivity and that the model does a reasonably good job of capturing it. In judging the performance of the model in this regard, it is useful to remember that we have gone beyond the qualitative prediction that subadditivity will be present in the data, to specifying both the amount of the effect and the conditions under which it will *not* be present. Given these goals, we view the results as supporting our model.

Individual Analyses

Because each subject rated all stimuli, we can fit models to the data of each individual. These results are shown in Table 3, where we have ordered subjects by their estimated θ's. The data of 5 subjects (Numbers 28–32 in Table 3) do not fit our theory. We rejected model fits for these subjects on either one of two grounds; namely, that the estimation procedure yielded theoretically impossible values (e.g., $\hat{\theta}' = -.16$ for Subject 28), or that the underlying data indicated a pattern of deviations from p_A that ran contrary to our theory. For example, Subject 31 generally adjusted downward for low probabilities but upward for higher values. For the majority of 27 subjects, however, the data accorded well with our theory. This can be observed by noting the substantial t statistics associated with the estimates of θ'. These are particularly critical to the theory because if $\theta' = 0$, there is no adjustment. On the other hand, the fact that $\hat{\beta}$ is not significantly greater than zero for some 15 out of 27 models does not violate the model because the theory defines $\beta \geq 0$. This does mean, however, that in this task, β takes on a low value for most subjects. (Recall $\hat{\beta} = .12$ for the aggregate data.)

Table 3 indicates substantial individual differences in the parameter values and in the degree to which subjects' models fit their data (as indicated by the MADs). When compared with the aggregate analysis, the individual models contain considerably more noise (the MAD for the aggregate data is .021). In Table 4 we illustrate individual differences further by presenting the results of 5 subjects, each representing a different combination of θ' and β parameters. Subject 21 illustrates the use of a highly consistent strategy in which downward adjustments are made over almost the entire range of p. Subject 17 also has a consistent strategy involving adjustments, but $\hat{p}_c = .49$, implying that adjustments will be down when $p_A > .49$, up when $p_A < .49$, and have no adjustments at $p_A = .49$. The data conform quite closely to this pattern. Subject 12 has a somewhat less consistent strategy of making small upward adjustments over most of the range of p ($\hat{p}_c = .83$). Again, the data are generally consistent with this interpretation. Subject 1 is included for contrast because, as can be seen, there was almost total reliance on p_A (as would be predicted by the parameter values and low MAD). Subject 27 is shown to illustrate the most extreme and least consistent adjustment process (which was generally downward). As is evident from the data, this subject had difficulty in controlling the adjustment process (cf. Hammond & Summers, 1972, on cognitive control). This lack of consistency manifested itself in widely different adjust-

Table 2
Subadditivity for the Aggregate Data

p_A	$(1 - p_A)$	n	Actual: $\hat{S}(f:c) + \hat{S}(c:f)$	Predicted: $\hat{S}(f:c) + \hat{S}(c:f)$
1	0	2	1.01	1.00
1	0	6	0.99	1.00
1	0	12	1.01	1.00
1	0	20	0.99	1.00
.89	.11	9	1.00	.97
.89	.11	18	1.00	.98
(.89)	(.11)	(18)	(0.98)	(.98)
.80	.20	5	1.01	.94
.80	.20	10	0.92	.97
(.80)	(.20)	(10)	(0.97)	(.97)
.75	.25	4	0.83	.92
.67	.33	6	0.92	.95
(.67)	(.33)	(6)	(0.92)	(.95)
.67	.33	9	0.88	.96
.67	.33	18	0.92	.99
.60	.40	5	0.89	.94
.60	.40	10	0.97	.97
.50	.50	2	0.90	.84
.50	.50	8	0.88	.96
(.50)	(.50)	(8)	(0.94)	(.96)
.50	.50	12	0.95	.98
.50	.50	20	0.94	.98

Note. Numbers in parentheses are for the repeat judgments. p_A = initial assessment.

ments for the same stimuli as well as illogical judgments. An example of the latter was that evidence of (0:2) was evaluated as *stronger* than (2:0) (i.e., .40 vs. .30). The lack of consistency and large amount of adjusting that characterize Subject 27 suggested that there might be a positive relation between the size of θ' and MAD, over subjects. When we investigated this for Subjects 1–27, the correlation was $r = .73$ ($p < .001$). Thus, there seems to be a connection between the amount of adjustment and the ability to execute it consistently.

Our final results concern the additivity/nonadditivity of complementary probabilities for individual subjects. This is illustrated in Table 5, using the subjects discussed. The important thing to note is that Subject 21 is consistently subadditive (and this is predicted quite well by the model); Subject 17 is generally additive, as implied by $\hat{\beta} = .95$; Subject 12 is superadditive, but not consistently so; Subject 1 is additive; and Subject 27 is both highly subadditive and inconsistent. From our perspective, these results strengthen our interpretation of the θ and β parameters, as well as the general form of the model.

A possible criticism of the above experiment is that although we investigated the responses of 32 individual subjects in depth, we only obtained responses to a single scenario. In other words, were our results simply a function of the content of the specific scenario investigated? We ran, therefore, another 32 subjects, using four different content scenarios but with

Table 3
Fit of the Model for Individual Subjects

Subject	$\hat{\theta}'$	(t)	$\hat{\beta}$	(t)	$\hat{p}_c$	MAD
1	0.04	(4.0)	0.01	(0.17)	.03	.006
2	0.06	(2.9)	0.00	(0.00)	.00	.015
3	0.09	(3.0)	0.41	(0.67)	.35	.042
4	0.10	(2.5)	0.19	(0.43)	.24	.062
5	0.12	(12.0)	0.02	(0.89)	.05	.009
6	0.17	(4.2)	0.16	(0.84)	.22	.052
7	0.17	(5.7)	0.09	(1.5)	.15	.026
8	0.22	(11.0)	0.23	(1.4)	.26	.032
9	0.25	(6.3)	18,771.00	(∞)	.99	.051
10	0.26	(8.7)	0.01	(0.70)	.02	.038
11	0.36	(7.2)	0.01	(0.76)	.02	.071
12	0.37	(10.0)	9.40	(1.17)	.83	.054
13	0.38	(12.7)	0.11	(2.8)	.17	.033
14	0.41	(13.7)	0.05	(2.3)	.10	.028
15	0.43	(10.8)	0.01	(0.90)	.02	.053
16	0.44	(8.6)	0.06	(2.1)	.12	.080
17	0.44	(14.7)	0.95	(5.3)	.49	.030
18	0.46	(11.5)	0.15	(2.1)	.21	.059
19	0.46	(9.2)	0.02	(1.5)	.03	.078
20	0.49	(12.2)	0.03	(3.0)	.07	.047
21	0.50	(16.7)	0.02	(2.9)	.05	.023
22	0.70	(14.0)	0.01	(1.0)	.03	.089
23	0.73	(14.6)	1.51	(6.0)	.57	.071
24	0.80	(20.0)	0.08	(4.7)	.14	.053
25	0.90	(22.5)	0.64	(6.4)	.42	.070
26	1.24	(22.5)	0.02	(3.0)	.06	.089
27	1.68	(28.0)	0.05	(0.71)	.10	.108
28	−0.16		0.01		.02	.052
29	0.03		−0.04		—	.025
30	0.05		−0.01		—	.112
31	0.08		15,052.00		.99	.052
32	0.35		−0.01		—	.079

Note. $\hat{\theta}'$ = estimated amount of ambiguity; (t) = t-test statistic; $\hat{\beta}$ = estimated attitude toward ambiguity; $\hat{p}_c$ = estimated crossover point; MAD = mean absolute deviation.

the same numerical values as in the scenario involving the automobile accident. These scenarios involved (a) a taste test where people had to identify a soft drink (Coke vs. Pepsi), (b) a bank robbery where witnesses said the robbers spoke to each other in a foreign language (German vs. Italian), (c) an experiment where 6-year-old children had to identify words flashed on a screen (ROT vs. BED), and (d) experts investigating the cause of a fire (arson vs. short circuit). Eight subjects were assigned at random to each scenario. Because the results from these four scenarios parallel those of the automobile-accident scenario in terms of model fits (albeit with different parameter values) they are not presented here.

Experiment 2

We had two goals in conducting Experiment 2. First, we wished to test systematically for the effects of source credibility and signal dissimilarity on the parameters of the model. In accordance with our theory, θ' should decrease as source credibility and signal dissimilarity increase. Second, we wished to investigate the importance of individual differences in the way people cope with the ambiguity inherent in our judgment task.

Method

Design. Two levels (high/low) of source credibility and dissimilarity of signals were crossed in a 2 × 2 factorial design. In addition, four different content scenarios were constructed that varied on all four experimental combinations (resulting in 16 different stories). Subjects were asked to judge 21 stimuli that varied in p and n (see below) for each of the four content-distinct scenarios. Thus, each subject initially made 84 probability judgments. However, to reduce boredom in the task, subjects made judgments in all four scenarios, with each scenario representing one of the four experimental conditions. For example, Subject 1 received Scenario A in the high/high condition, Scenario B in the high/low condition, and so on. A four-person Latin square was set up so that every scenario appeared an equal number of times in each experimental condition. Finally, because subjects made judgments in one scenario under the high/high condition, the same scenario was also given in the low/low condition (and the order was counterbalanced). In this way, we were able to examine each subject's judgments, holding the content of the scenario constant. This part of the experiment required 21 additional judgments, making the total number of responses for each subject equal to 105.

Stimuli. The four content scenarios used involved the automobile accident from Experiment 1, the word-recognition task described above, and two new stories. These latter scenarios involved determining the name of a play from an excerpt and the diagnosis of a medical condition. Four versions of each scenario were constructed to reflect different levels of credibility and dissimilarity (e.g., in the word-recognition task, we had 15-year-olds vs. 6-year-olds and BED vs. ROT as opposed to BED vs. BID). Within each scenario, subjects were given 21 stimuli that reflected the amount of evidence for each hypothesis. The values of the stimuli were different from those used in Experiment 1 in that smaller values of n were used in order to provide more sensitive tests of the model. The stimuli used were as follows: for $p = 0, 1, n = 1, 2, 6$; for $p = .125, .875, n = 8$; for $p = .2, .8, n = 5$; for $p = .25, .75, n = 4$; for $p = .33, .67, n = 6, 9$; for $p = .67, n = 3$; for $p = .4, .6, n = 5$; and for $p = .5, n = 2, 8$.

Subjects and procedures. Thirty-two subjects participated in this experiment (constituting eight 4-person Latin squares). Subjects were paid $5 per hour, and the task took about one hour to complete. The tasks were presented in booklets, and after each series of 21 judgments, subjects were either given a break or another task. At the end of the experiment, a manipulation check was performed on the credibility and dissimilarity induction. Specifically, each subject was asked to rate (using a 0–100 scale) the credibility of the source and the confusability of the signals in all four scenarios. Because each scenario had high and low levels of each factor, the subjects rated credibility and dissimilarity under both conditions. Therefore, subjects made four judgments on each of the four scenarios.

Results

Before presenting the main results, we note that the manipulation check showed that subjects did, on average, see the high-credibility versions of the same scenarios as greater than the low (80 vs. 47), and the high-dissimilarity signals as less confusable than the low (30 vs. 62).

General fit of the model. For each subject in each experimental condition, the model was fit to yield estimates of θ' and β (this resulted in 160 models–32 subjects × 5 models). The fit of the individual models was comparable to that of Experiment 1 (median MAD = .042, over all conditions).

Manipulation of θ'. The appropriate analysis of variance (ANOVA; 2 × 2 × 4) was performed using $\hat{\theta}'$ as the dependent variable. The results showed a significant main effect for *credibility* ($p < .001$), but no main effect for *dissimilarity*, and no significant interactions. The results for the main effect are shown in Figure 4. Note that θ' increases as the credibility of the source decreases, thereby confirming our prediction. However, contrary to our prediction, there was no effect for dissimilarity.

Table 4
Fit of the Model for Selected Subjects

		Subject									
		21		17		12		1		27	
n	p_A	S	Ŝ	S	Ŝ	S	Ŝ	S	Ŝ	S	Ŝ
2	1	.80	.75	.70	.78	.92	.92	.99	.98	.30	.16
6	1	.89	.92	.89	.93	.46	.94	.99	.99	.80	.72
12	1	.97	.96	.90	.96	.95	.97	.99	1.00	.80	.86
20	1	.98	.98	.88	.98	.88	.98	.99	1.00	.70	.92
9	.89	.83	.84	.84	.85	.84	.88	.88	.89	.90	.72
18	.89	.84	.87	.85	.87	.87	.89	.89	.89	.80	.81
(18)	.89	.87	.87	.85	.87	.87	.89	.89	.89	.70	.81
27	.89	.85	.87	.82	.88	.82	.89	.89	.89	.60	.83
5	.80	.77	.72	.80	.75	.85	.81	.80	.79	.60	.53
(10)	.80	.72	.76	.80	.77	.83	.80	.80	.80	.90	.67
10	.80	.76	.76	.80	.77	.84	.80	.80	.80	.60	.67
15	.80	.79	.77	.79	.78	.74	.80	.80	.80	.70	.71
20	.80	.76	.78	.79	.79	.72	.80	.80	.80	.60	.73
25	.80	.76	.78	.80	.79	.85	.80	.80	.80	.90	.75
(25)	.80	.76	.78	.80	.79	.74	.80	.80	.80	.70	.75
4	.75	.57	.66	.65	.69	.72	.77	.66	.74	.30	.44
3	.67	.57	.56	.64	.62	.74	.71	.66	.66	.20	.31
(3)	.67	.61	.56	.65	.62	.75	.71	.66	.66	.20	.31
6	.67	.59	.61	.65	.64	.72	.69	.67	.66	.60	.49
(6)	.67	.57	.61	.65	.64	.72	.69	.66	.66	.30	.49
9	.67	.59	.63	.65	.65	.65	.68	.66	.67	.30	.55
12	.67	.60	.64	.64	.66	.72	.68	.66	.67	.40	.58
15	.67	.62	.65	.65	.66	.73	.68	.66	.67	.70	.60
18	.67	.63	.65	.65	.66	.65	.68	.67	.67	.70	.61
24	.67	.62	.66	.65	.66	.63	.67	.66	.67	.50	.62
5	.60	.52	.54	.60	.58	.67	.63	.60	.59	.60	.41
10	.60	.57	.57	.60	.59	.66	.61	.60	.60	.40	.50
2	.50	.38	.38	.50	.50	.57	.59	.50	.49	.20	.11
8	.50	.42	.47	.49	.50	.53	.52	.50	.50	.30	.40
(8)	.50	.50	.47	.50	.50	.52	.52	.50	.50	.30	.40
12	.50	.47	.48	.50	.50	.54	.51	.50	.50	.30	.43
20	.50	.48	.49	.51	.50	.55	.51	.50	.50	.30	.46
5	.40	.36	.36	.40	.42	.42	.44	.40	.40	.20	.28
10	.40	.40	.38	.40	.41	.34	.42	.39	.40	.50	.34
6	.33	.27	.30	.31	.35	.25	.37	.34	.33	.40	.25
(6)	.33	.30	.30	.34	.35	.38	.37	.34	.33	.20	.25
9	.33	.26	.31	.34	.36	.35	.36	.33	.33	.20	.28
18	.33	.30	.32	.34	.34	.33	.34	.33	.33	.30	.30
4	.25	.25	.22	.25	.30	.24	.32	.25	.25	.10	.17
5	.20	.18	.18	.21	.25	.23	.26	.20	.20	.20	.16
10	.20	.18	.19	.21	.23	.22	.23	.20	.20	.10	.18
(10)	.20	.20	.19	.40	.23	.26	.23	.20	.20	.20	.18
9	.11	.08	.11	.17	.15	.12	.15	.11	.11	.10	.11
18	.11	.08	.11	.15	.13	.12	.13	.12	.11	.10	.11
2	0	.05	.25	.10	.22	.14	.08	.00	.02	.40	.84
6	0	.03	.08	.10	.07	.13	.06	.00	.01	.30	.28

Table 4 (*continued*)

		Subject									
		21		17		12		1		27	
n	p_A	S	Ŝ	S	Ŝ	S	Ŝ	S	Ŝ	S	Ŝ
12	0	.02	.04	.10	.04	.14	.03	.00	.00	.20	.14
20	0	.02	.03	.11	.02	.12	.02	.00	.00	.10	.08
			$\hat{\theta}' = 0.50$		$\hat{\theta}' = 0.44$		$\hat{\theta}' = 0.37$		$\hat{\theta}' = 0.04$		$\hat{\theta}' = 1.68$
			$\hat{\beta} = 0.02$		$\hat{\beta} = 0.95$		$\hat{\beta} = 9.4$		$\hat{\beta} = 0.01$		$\hat{\beta} = 0.05$
			$\hat{p}_c = .05$		$\hat{p}_c = .49$		$\hat{p}_c = .83$		$\hat{p}_c = .03$		$\hat{p}_c = .10$
			MAD = .023		MAD = .030		MAD = .054		MAD = .006		MAD = .108

Note. p_A = initial assessment; S = subject response; Ŝ = predicted response; $\hat{\theta}'$ = estimated amount of ambiguity; $\hat{\beta}$ = estimated attitude toward ambiguity; $\hat{p}_c$ = estimated crossover point; MAD = mean absolute deviation.

In addition to the above analysis, recall that each subject also received the *same* scenario in the high/high and low/low conditions. A comparison of the means of the estimated θ's in these two conditions also showed a significant difference in the hypothesized direction;

Table 5
Additivity/Nonadditivity of Complementary Probabilities

			Subject									
			21		17		12		1		27	
p_A	$(1 - p_A)$	n	Σ	$\hat{\Sigma}$	Σ	$\hat{\Sigma}$	Σ	$\hat{\Sigma}$	Σ	$\hat{\Sigma}$	Σ	$\hat{\Sigma}$
1.00	.00	2	0.85	1.00	0.80	1.00	1.06	1.00	0.99	1.00	0.70	1.00
1.00	.00	6	0.92	1.00	0.99	1.00	0.59	1.00	0.99	1.00	1.10	1.00
1.00	.00	12	0.99	1.00	1.00	1.00	1.09	1.00	0.99	1.00	1.00	1.00
1.00	.00	20	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	0.80	1.00
.89	.11	9	0.91	0.95	1.01	1.00	0.96	1.03	0.99	1.00	0.80	0.83
.89	.11	18	0.92	0.97	1.00	1.00	0.99	1.02	1.01	1.00	0.90	0.92
.89	.11	18	0.95	0.97	1.00	1.00	0.96	1.02	1.01	1.00	0.80	0.92
.80	.20	5	0.95	0.90	1.01	1.00	1.08	1.07	1.00	0.99	0.80	0.69
.80	.20	10	0.94	0.95	1.01	1.00	1.06	1.03	1.00	1.00	0.70	0.85
.80	.20	10	0.96	0.95	1.20	1.00	1.10	1.03	1.00	1.00	0.80	0.85
.75	.25	4	0.82	0.88	0.90	0.99	0.96	1.09	0.91	0.99	0.40	0.61
.67	.33	6	0.86	0.92	0.99	0.99	0.97	1.06	1.01	0.99	1.00	0.74
.67	.33	6	0.87	0.92	0.96	0.99	1.10	1.06	1.00	0.99	0.50	0.74
.67	.33	9	0.85	0.95	0.99	1.00	1.00	1.04	0.99	1.00	0.50	0.83
.67	.33	18	0.93	0.97	0.99	1.00	0.98	1.02	1.00	1.00	1.00	0.91
.60	.40	5	0.88	0.90	1.00	1.00	1.09	1.07	0.99	0.99	0.80	0.69
.60	.40	10	0.97	0.95	1.00	1.00	1.00	1.03	1.00	1.00	0.90	0.84
.50	.50	2	0.76	0.75	1.00	1.00	1.14	1.18	1.00	.98	0.40	0.22
.50	.50	8	1.00	0.94	0.98	1.00	1.06	1.04	1.00	1.00	0.60	0.80
.50	.50	8	0.84	0.94	1.00	1.00	1.04	1.04	1.00	1.00	0.60	0.80
.50	.50	12	0.94	0.96	1.00	1.00	1.08	1.02	1.00	1.00	0.60	0.86
.50	.50	20	0.96	0.98	1.02	1.00	1.10	1.02	1.00	1.00	0.60	0.92
				$\hat{\theta}' = 0.50$		$\hat{\theta}' = 0.44$		$\hat{\theta}' = 0.37$		$\hat{\theta}' = 0.04$		$\hat{\theta}' = 1.68$
				$\hat{\beta} = 0.02$		$\hat{\beta} = 0.95$		$\hat{\beta} = 9.40$		$\hat{\beta} = 0.01$		$\hat{\beta} = 0.05$

Note. p_A = initial assessment; Σ = sum of complementary probabilities; $\hat{\Sigma}$ = predicted sum of complementary probabilities; $\hat{\theta}'$ = estimated amount of ambiquity; $\hat{\beta}$ = estimated attitude toward ambiguity.

		Dissimilarity: High	Dissimilarity: Low	
Credibility	High	.20	.22	.21
	Low	.35	.32	.33
		.28	.27	

Figure 4. Mean θ' parameters by experimental conditions.

that is, $\bar{\theta}' = .19$ in the high/high condition, $\bar{\theta}' = .32$ in the low/low ($p < .005$, by a paired t test). Thus, with the exception of an effect for the dissimilarity of the signals, our hypotheses concerning θ' are supported by the experimental data.

Individual differences. We now consider the following: (a) Can subjects be characterized as having a general strategy, as measured by the consistency of their θ' and β values, in different scenarios? (b) Is the amount of one's adjustment, as measured by θ', systematically related to the consistency of executing one's strategy? (c) Can individual perceptions of the credibility of the source and the dissimilarity of the signals account for variance in θ' and β *within* each of the experimental conditions?

1. Recall that for each subject, four different scenarios were given and a model fit to the data in each. Therefore, each subject can be characterized by four θ's, βs, and MADs. To determine if the parameter values were more alike *within* a subject than between subjects (this is measured by the intraclass correlation), a one-way repeated ANOVA was performed (32×4) for $\hat{\theta}'$, $\hat{p}_c$, and MAD (Winer, 1963, chap. 3). The results show that for $\hat{\theta}'$, $r = .76$ ($p < .001$); for $\hat{p}_c$, $r = .72$ ($p < .001$); and for MAD, $r = .81$ ($p < .001$). These results are particularly striking when one recalls that the four scenarios varied over the four experimental conditions. However, in spite of these differences, the results show strong and stable individual strategies in the amount that is adjusted (θ'), the direction of the adjustments (p_c or β), and the consistency of executing one's strategy (MAD).

2. In Experiment 1, we found a significant positive correlation between $\hat{\theta}'$ and MAD. The same positive relation was found here in all four scenarios ($r = .75, .76, .46, .43$). Thus, our interpretation of θ as reflecting a cognitive simulation process is strengthened by the generality of this finding.

3. Because each subject made independent judgments of the credibility and confusability of the experimental stimuli, we were also able to investigate how these judgments related to θ' *within* experimental conditions. To do so, we reanalyzed our data with a regression model in which θ' was the dependent variable and the individual ratings of credibility and confusability, together with dummy variables representing the different scenarios, were the independent variables. More precisely, there is a regression equation of this type for each of the four experimental conditions. However, these four equations can be estimated more efficiently as a single model using Zellner's (1962) procedure for "seemingly unrelated" regressions. The R estimated by this procedure was .49 (with an adjusted R of .40). Of the independent variables, there was no effect for either scenarios or confusability. However, all four coefficients for credibility in the different experimental conditions were significant ($p < .005$) and of the hypothesized sign (i.e., a negative relation between θ' and ratings of credibility). We interpret these results as strengthening the conclusions drawn from the more standard ANOVA of our study; that is, θ' is not only affected by different levels of credibility across all subjects, it also covaries significantly with individual perceptions of credibility within each of these levels.

Experiment 3

The purpose of this experiment was to answer the following question: Can individuals' choices between gambles be predicted from knowledge of their θ' and β parameters obtained from a separate inference task? To examine this, subjects were first asked to make judgments in Experiments 1 and 2 and both θ' and β were estimated as before. The subjects were then asked to choose (or express indifference) between nine pairs of gambles involving the outcome from an urn with known probability versus the occurrence of an event on the basis of unreliable reports. If θ' and β do capture aspects of ambiguity that affect choice, knowledge of these parameters should allow one to predict individual choices in addition to inferences.

Method

Subjects. Twenty subjects, recruited from the University of Chicago community, participated in this study. They were paid \$5 an hour.

Stimuli. For the inference task, two different scenarios were used: the automobile-accident story and the taste-test story (Pepsi vs Coke), for which we had also previously collected data (see end of Experiment 1). These were chosen because the θ and β values were quite different in the two cases. In both scenarios, subjects received 48 stimuli. These were the same 40 combinations of p and n, together with eight repeats, used in Experiment 1. The stimuli for the choice task involved one of the following: (a) In the automobile-accident task, subjects were faced with choosing between betting that a ball drawn from an urn with known probability was green, versus betting that the car that caused the accident was green, based on witnesses' reports of the car color. (b) For those in the taste-test scenario, the choice was similarly between betting that the outcome from an urn was a certain color, versus betting that the drink was Pepsi or Coke. In both scenarios, subjects were told to imagine that their payoff for being correct would be \$10. Thus, the payoffs for the urn gamble and the bet involving the report of some event were equal. Within scenarios, each subject saw nine pairs of gambles that varied in the proportion of colored balls in the urn and the proportion of reports favoring the particular hypothesis. These proportions were always the same in the two bets. The exact values of p used in the nine pairs were 1, .875, .75, .625, .50, .375, .25, .125, and 0. The number of balls in the urn and the number of reports were held constant at eight.

Procedure. The 20 subjects were randomly assigned to one of the two scenarios. The procedure for the inference task was identical to the previous experiments. After subjects finished the inference task, they were presented with the appropriate choice task. The nature of the two gambles was explained, and subjects were then asked to choose or to indicate indifference between the gambles. If they were not indifferent, they were also asked to indicate their strength of preference on a 5-point scale (from *little* to *great deal*). After doing this for one value of p, they turned the page and made another choice (and strength of preference rating, if appropriate) at the next level of p. This continued until all nine pairs had been considered. Therefore, for each subject, there were nine choices between an unambiguous bet from an urn with known p, versus an ambiguous bet that an event occurred, on the basis of the proportion of favorable reports from an unreliable source.

Results

Because each subject first participated in the inference task, we briefly consider these results before discussing the choice data. The estimated θ's and p_cs (implied by βs) are shown in Table 6. As expected, there were marked differences in the θ' and β parameters in the two scenarios. On average, the automobile-accident scenario induced less adjustment and a lower crossover point than did the taste-test scenario. In fact, the first 2 subjects in the accident scenario had nonsignificant θ' values (i.e., $\theta' = 0$). For two other subjects (Numbers 9 and 10), the estimation procedure yielded parameter values inconsistent with our theory. Because these subjects did not fit the model, we made no predictions concerning their choices.

To compare each subject's choices with predictions from the inference model, the following procedure was used: Any combination of θ' and p_c implies when and where $S(p_A)$ is greater, less than, or equal to, p_A (see Equation 8). Thus, for each subject, when $p_A > S(p_A)$, we predicted the urn would be chosen over the bet based on unreliable reports; when $S(p_A) = p_A$, we predicted indifference between the two gambles. For the 2 subjects (Numbers 1 and 2) whose estimated θ' parameters were not significantly different from zero, we always predicted indifference between the gambles because $\theta' = 0$ implies that $S(p_A) = p_A$ for all p_A.

Table 6
Choice Predictions From Knowledge of $\hat{\theta}'$ and $\hat{p}_c$

Subject	$\hat{\theta}'$	$\hat{p}_c$	No. of hits	Probability ($r \geq$ hits)
	Automobile-accident scenario			
1	0.02	—	8	.001
2	0.03	—	4	.341
3	0.03	.51	6	.019[a]
4	0.10	.15	6	.040
5	0.25	.37	6	.040
6	0.42	.32	5	.140
7	0.60	.03	5	.140
8	0.72	.47	6	.040
9	−0.05	—	—	—
10	−0.36	—	—	—
	Taste-test scenario			
11	0.17	.36	5	.140
12	0.23	.34	8	.001
13	0.77	.47	1	.849
14	0.80	.50	4	.341
15	0.84	.45	6	.040
16	1.19	.49	7	.008
17	1.38	.39	6	.040
18	1.54	.50	5	.140
19	1.61	.37	7	.008
20	2.64	.05	3	.612

Note. $\hat{\theta}'$ = estimated amount of ambiguity; $\hat{p}_c$ = estimated crossover point.
[a] Subject made only eight choices.

The fourth column of Table 6 shows the number of correct choice predictions by subject.

To evaluate how well the choices were predicted from knowledge of θ' and $\hat{p}_c$, we used a random baseline for comparison; that is, for each of the nine choices made by a subject, there were three possible outcomes: urn, report, and indifference. Because the probability of randomly predicting the correct response is 1:3, we computed the probability of getting at least r hits in nine trials on the basis of chance (using the binomial distribution).[4] This probability is shown in the last column of Table 6. For example, Subject 1 was correctly predicted in eight of the nine choices; the probability of getting at least this many hits by chance is .001. Thus, we rejected the hypothesis that our predictions for this subject were no better than chance. Using this method for all subjects, it can be seen that 5 of the 8 subjects for whom we made predictions in the automobile-accident scenario, and 5 of 10 in the taste test, are well predicted using a Type 1 error level of .05. If this error level were increased to .15, a majority of subjects (14:18) would be accurately predicted from their inference parameters. In any event, at the aggregate level (over subjects and scenarios), there were 98 hits out of 161 predictions (one response was missing). The probability of getting at least this many hits by chance is miniscule.

Second, consider the results concerning the strength of preference ratings. Recall that in addition to choosing between gambles, subjects were asked to rate their strength of preference on a 5-point scale. These ratings supplement our analysis of the choice data in the following way: In the taste test, θ' is much larger than in the automobile-accident scenario. Because θ' is directly related to the amount of adjustment to p_A, the differences between $S(p_A)$ and p_A should be larger in the taste test than in the accident story. Furthermore, the larger the differences, the stronger one's preferences should be because they are further away from indifference (where $p_A = S(p_A)$). We tested this by comparing the mean strength-of-preference ratings in the two stories across the nine levels of p. These results are shown in Table 7. First, note that the means for the taste test are larger than the means for the automobile accident at every level of p. Second, the pattern of means is consistent with the general form of the model

Table 7
Means of Strength-of-Preference for Two Scenarios

p	Automobile accident	Taste test
1.000	2.9	3.7
.875	2.3	3.1
.750	1.8	2.5
.625	1.6	2.1
.500	1.0	1.7
.375	1.3	2.1
.250	1.3	2.1
.125	1.1	2.0
.000	2.1	2.2
Grand Mean	1.71	2.39

in that preferences are strongest at $p = 1$, decrease as p approaches p_c, and then increase again at $p = 0$. Therefore, the strength-of-preference data are consistent with both the difference in the sizes of θ' for the two scenarios as well as for the general form of the model.

As the astute reader may have noticed, our theory does not necessarily imply exact equivalence between choice and inference tasks because these could differ with respect to the β parameter. In particular, whereas payoffs are explicit in the choice task (i.e., a gain of $10), there are no explicit payoffs in the inference task. Thus, one might expect a systematic bias between β as estimated in the inference task, and β as implied by subjects' choices. Specifically, as stated after first presenting our model, if the effect of ambiguity is to induce caution rather than to riskiness, then the prospect of a gain would focus attention more on smaller rather than on larger values of p(gain), such that $\beta_{\text{choice}} < \beta_{\text{inference}}$. (Conversely, the prospect of a loss would imply more attention being paid to greater rather than smaller values of p(loss), such that $\beta_{\text{choice}} > \beta_{\text{inference}}$.) Consequently, one would expect ambiguity avoidance over a wider range of p in tasks involving choice as opposed to inference. Indeed, some of the errors in predicting subjects' choices can be attributed to precisely this source of sys-

[4] Although it can be argued that the indifference response is not as likely as are the other two responses, it must be remembered that in experiments involving ambiguity, the indifference response *is* quite prevalent. Indeed, in our own work with the Ellsberg paradox, indifference responses occurred approximately ⅓ of the time.

tematic bias. Consider the data for the automobile-accident scenario in Table 6. Several subjects had low $\hat{p}_c$ values in the inference task and thus could be thought of as having already conceptualized the task in terms of "gains." On the other hand, for the 4 subjects with relatively high $\hat{p}_c$s in the inference task (Numbers 3, 5, 6, and 8), 10 out of 12 prediction errors were in the same direction; namely, subjects' $\hat{\beta}$s estimated in the inference task indicated larger $\hat{p}_c$s than were revealed by their choices. The same bias was also found in the taste-test scenario. That is, consider again those subjects with relatively high $\hat{p}_c$s (i.e., all except Number 20). Excluding Subject 13, 17 out of 24 prediction errors are consistent with the $\beta_{\text{choice}} < \beta_{\text{inference}}$ bias. The eight prediction errors of Subject 13 are distributed in both directions. To summarize, we conclude that whereas individuals' parameters in an inference task can be used to predict choices, many errors of prediction are in accord with a systematic bias in the β parameter that is consistent with our theory.

Experiment 4

Having manipulated the θ parameter in Experiment 2, we designed Experiment 4 to investigate the effects of manipulating β. This was done by allocating subjects to different roles (sellers and buyers) in an insurance context. The dependent variable of interest involved statements of maximum buying prices and minimum selling prices. The data were collected as part of a larger investigation by Hogarth and Kunreuther (1984) on the effects of ambiguity in insurance decision making.

The assumption underlying the experimental manipulation is that a person who assumes a risk, is likely to pay more attention to larger values of p(loss) than is someone who transfers the risk. Experimental evidence consistent with this assertion has been documented by Hershey, Kunreuther, and Schoemaker (1982) and Thaler (1980). In our framework, it implies that $\beta_{\text{seller}} > \beta_{\text{buyer}}$. Given this assumption, approximate ambiguity functions for buyers and sellers of insurance can be sketched as in Figure 5. Note that when buyers/sellers are in a nonambiguous situation, $S(p_A) = p_A$, and all responses are on the diagonal.

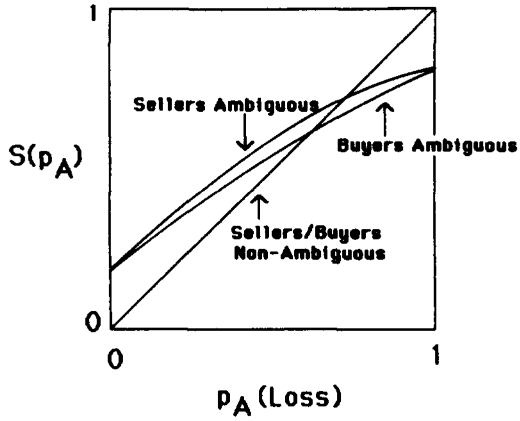

Figure 5. Approximate ambiguity functions for buyers and sellers of insurance.

If one further assumes that buying and selling prices are the same monotonic function of $S(p_A)$, Figure 5 suggests the following predictions: (a) When buyers and sellers are in a nonambiguous situation, $S(p_A) = p_A$, and the seller's price should equal the buyer's. (b) When buyers and sellers are equally ambiguous (i.e., their θs are equal), the seller's price should exceed the buyer's over the whole range of p_A. Note that this arises because the seller always weights larger imaginary values of p(loss) more than does the buyer. (c) When a seller has no ambiguity about the probability of a loss, but a buyer does, the buyer will perceive the probability of loss as higher than the seller, and should be willing to pay *more* than the seller would ask. However, when $p_A > p_c$, the buyer will perceive the loss probability as lower than the seller, and offer less than the seller would ask. In Figure 4, this is shown by comparing the buyer-ambiguous function with the diagonal (seller unambiguous). Note that the buyer's function is above the diagonal for $p_A < p_c$, and lower than the diagonal for $p_A > p_c$. This implication of the model provides a particularly stringent test for our theory. Experiment 4 was designed to test the above three predictions.

Method

Design. Prices for insurance contingent on ambiguous and nonambiguous probabilities were investigated across four different probability levels (.01, .35, .65, and .90). Each subject was assigned the role of buyer or seller of a contract concerning a potential $100,000 loss and responded to both ambiguous and nonambiguous versions of the stimulus at one probability level. Thus, the design of the experiment involved three factors, two of which were *between* subjects (i.e., role of buyer or seller, and probability level) and one *within* subjects (i.e., ambiguous vs. nonambiguous probabilities).

Stimuli. The scenario used in the stimulus material involved the owner of a small business (net assets of $110,000) who was seeking to insure against a $100,000 loss that could result from claims concerning a defective product. Subjects assigned the role of buyers were told to imagine they were the owner of the business. Subjects assigned the role of sellers were asked to imagine that they headed a department in a large insurance company and were authorized to set premiums for the level of risk involved. Ambiguity was manipulated by factors involving how well the manufacturing process was understood, whether the reliabilities of machines used in the process were known, and the extent to which manufacturing records were well kept. In both ambiguous and nonambiguous cases a specific probability level was stated (e.g., .01); however, a comment was added as to whether one could "feel confident" (nonambiguous case) or "experience considerable uncertainty" (ambiguous case) concerning the estimate. As far as possible, the same wording was used in both the buyer and seller versions so that perceptions of ambiguity would be uniform in the two cases.

Subjects and procedures. Subjects were 112 MBA (masters of business administration) students at the University of Chicago who responded to questionnaires distributed in a course on decision making. To avoid prior influence, the experiment took place during the beginning of classes. Subjects were asked to respond to the questionnaire in anonymous fashion and were promised group-level feedback at a later class session (which they subsequently received). Note that subjects had prior training in business, economics, and statistics, and the insurance context was familiar to them. Eight different forms of the stimulus materials, corresponding to the 2 (roles) × 4 (probability levels), were shuffled and distributed in the classrooms, thereby ensuring random allocation of subjects to conditions. After reading each stimulus, subjects were asked to state maximum buying prices (for buyers) or minimum selling prices (for sellers).

Results

Table 8 shows the medians for all experimental conditions. We report medians because several distributions within conditions are quite skewed, and variances also differ between cells at the same probability levels, often significantly. The pattern of results in Table 8 supports our three predictions. First, in comparing buyers and sellers in the nonambiguous case, columns 1 versus 3, note that the median prices are quite similar over the four probability levels. Second, when buyers and sellers are both ambiguous, columns 2 versus 4, observe that the selling price is considerably larger than the buying price at every level of p. This result strongly confirms the notion that $\beta_{\text{seller}} > \beta_{\text{buyer}}$, when considering ambiguous loss probabilities. Third, consider the ambiguous buyer and the nonambiguous seller, columns 2 versus 3. As expected, when p is small (.01), the ambiguous buyer is willing to pay more ($1,500) than the nonambiguous seller asks ($1,000). However, as the probability of loss increases, the two prices converge (at p = .35), and then diverge, with the buyer's price being *less* than the seller's (at p = .65, .90). Indeed, at higher probabilities, the buyer is willing to spend considerably less than the seller wants to charge. Therefore, although the buyers seem to be ambiguity-avoiding at low probabilities, they paradoxically appear to be ambiguity-seeking for high-loss probabilities. Both results are predicted by our model. Although the data in Table 8 could be subjected to various statistical tests, the close match between the pattern of results and theoretical predictions makes this unnecessary. For example, in the comparison of columns 1 and 3, two of the four predictions are precisely correct (and the other two are not far off); the four predictions involving columns 2 and 4 are all correct, as are the predictions concerning columns 2 versus 3. Furthermore, the results within buyers and sellers (across ambiguity conditions) also conform to the model predictions. In Hogarth and Kunreuther (1984), the results of several other related experiments are reported using different scenarios, research designs, subjects, and response modes. The results of these experiments are consistent with those reported here, thus attesting to the stability of the phenomena.

Discussion

We now discuss the implications of our theory and results with respect to the following issues: (a) the importance of ambiguity in assessing uncertainty, (b) the use of cognitive strategies in probabilistic judgments under ambiguity, (c) the role of ambiguity in risky choice, and (d) extensions of the model to multiple sources and time periods.

Ambiguity and the Assessment of Uncertainty

The concept of ambiguity highlights the distinction between one's lack of knowledge

Table 8
Median Buying and Selling Prices for Insurance

	Buyers		Sellers	
Probability levels	(1) Nonambiguity	(2) Ambiguity	(3) Nonambiguity	(4) Ambiguity
.01	1,000	1,500	1,000	2,500
.35	35,000	35,000	37,500	52,500
.65	65,000	45,000	65,000	70,000
.90	82,500	60,000	90,000	90,000

Note. Numbers in columns 1–4 are in dollars.

of the process that generates outcomes and the uncertainty of outcomes *conditional* on some model of the process. The fact that there are at least two sources of uncertainty in most situations leads to the irony that one needs a well-defined model to give precise estimates of how much one doesn't know. Indeed, the usefulness of formulating well-defined stochastic processes is in eliminating ambiguity so that outcome uncertainty can be quantified. Thus, when coins are "fair" or random drawings are taken from urns with known *p*, there is no second-order uncertainty. Furthermore, the conditional nature of uncertainty is implicitly recognized in various attempts to quantify and improve inferential judgments. For example, consider how uncertainty is defined in the *lens model* (Hammond, Hursch, & Todd, 1964). In this case, the uncertainty in the environment is measured as the residual variance not accounted for by a well-formulated ecological model. Thus, unexplained variance or uncertainty is conditional on the model of how particular cues combine to form the criterion of interest. Now consider the work of Nisbett and colleagues on trying to improve probabilistic judgments through training (Nisbett, Kranz, Jepson, & Kunda, 1983; Jepson, Kranz, & Nisbett, 1983). They argue that training and experience can allow one to see the underlying structure of real-world problems so that the appropriate model can be used for making better judgments. Thus, the focus of their training is on making various statistical principles (e.g., regression to the mean, law of large numbers, use of base rates) more obvious in everyday inferences.

Although the conditional nature of uncertainty has been implicitly recognized, ambiguity results from its *explicit* recognition, that is, by realizing that the "model" is itself subject to uncertainty. Indeed, one could argue that the cost of urn models, coin-flipping analogies, and the like, is that they obscure the fact that most real-world generating processes are not precisely known. Furthermore, even if a process is well-defined at one point in time, the parameter(s) of the process can change over time, resulting in ambiguity as well as uncertainty. For example, imagine that you have been asked to evaluate the research output of a younger colleague being considered for promotion. Your colleague has produced 11 papers; of these the first 9 (in chronological order) represent competent, albeit unexciting, scholarly work. On the other hand, the latter two papers are quite different; they are innovative and suggest a creativity and depth of thought absent from the earlier work. What should you do? As someone who is aware of regression fallacies, you might consider the two outstanding papers as outliers from a stable generating process and thus predict regression to the mean. Alternatively, you might consider the outstanding papers as "extreme" responses that signal a change in the generating process, that is, a new and higher mean. If this were the case, the same general regression model would predict future papers of high quality (regression to a higher mean). If one asks what is the nature of the signaling in this case, it is obvious that the chronological order of the papers is crucial. Indeed, imagine that the outstanding papers were the first two that were written, or consider that they were the second and sixth. Each of these cases suggests a different underlying model and perhaps a different prediction.

Cognitive Strategies in Inferences Under Ambiguity

We have assumed that people use an anchoring-and-adjustment strategy in making inferences under ambiguity. However, whereas the term *anchoring-and-adjustment* is quite general and could encompass a wide range of models (cf. Lopes, 1981; Einhorn & Hogarth, in press), we have been quite specific as to the nature of this process in our tasks. Of greatest interest in this regard is the idea that adjustments are based on a mental simulation in which "what might be," or "what might have been," is combined with "what is" (the anchor). The rationale for this stems from the fact that the evaluation of evidence often involves an implicit comparison process (similar to the perception of figure against ground). Thus, when we evaluate the strength of evidence for a particular hypothesis, the evidence that might have been can serve as a convenient contrast case for comparison. Furthermore, because ambiguity implies that multiple models could have produced the observed results, it seems natural to consider that different results could have occurred on the basis of different underlying processes.

The support for the hypothesized anchoring-and-adjustment strategy comes from several sources. First, recall that in our model, the largest adjustments to the anchor occur when evidence is meager. Moreover, as n increases, $S(f{:}c)$ asymptotes at p_A. The results of Experiments 1 and 2 support this prediction. Thus, the weight of evidence (to use Keynes's term) for what is, dominates what might have been, as the absolute amount of evidence increases. Furthermore, the effect of increasing n is to reduce the amount of nonadditivity of complementary strengths. Because most of our subjects were subadditive, our model provides a psychological link to concerns expressed by others regarding the appropriateness of additivity when evidence is meager (Cohen, 1977; Shafer, 1976). In particular, Cohen (1977, chap. 3) pointed out that when one considers an incomplete system, the lower benchmark on provability is not necessarily disprovability, but nonprovability. For example, if one has meager circumstantial evidence such that the probability of the truth of a particular theory is .2, does this imply that the theory is false with $p = .8$? Rather, one might say that the theory is not proven (in a probabilistic sense), as opposed to saying that there is a .80 chance that it is wrong. Furthermore, the idea that the complement of statements can lead to *not proved,* rather than *disproved,* seems to be deeply imbedded in the Anglo-American legal system. Indeed, in Scottish law, defendants are either found guilty, not guilty, or *not proven.* The last category is reserved for those cases where the evidence is too meager to allow for a judgment of guilt or innocence.

Second, the fact that nonadditivity results from a shift in the direction of the adjustment process is consistent with other *order effects* due to the use of anchoring-and-adjustment strategies. For example, in a traditional Bayesian revision task, Lopes (1981) found that a change in the order in which sample information was presented affected overall judgments by changing the anchor. Thus, consider having to judge whether samples come from an urn containing predominantly red or blue balls (70/30 in both cases). You first draw a sample of eight that shows (5R:3B). Thereafter, you draw another sample of eight with the result (7R:1B). After each sample, you are asked how likely it is that you have drawn from the predominantly red urn. When the sample evidence is in the order given here, people seem to anchor on the first sample (5:3) and then adjust up for the second (stronger) sample. However, when the order of the samples is reversed, people anchor on (7:1) and adjust *down* for the weaker, second sample. This effect cannot be accounted for by assuming that people are using a Bayesian procedure (which treats the two situations as equal), but it does follow from an anchoring-and-adjustment process in which the anchor is weighted more heavily than the adjustment.

Third, the results of Experiment 2 provide important evidence regarding the process assumed to underlie the model. In addition to the fact that the experimental manipulation of source credibility affected θ' as predicted, two other results were found: a positive correlation between θ' and MAD, and the stability of individual differences in θ', β, and MAD across scenarios. The first result bears directly on the nature of the adjustment process inasmuch as it suggests a "cost" of engaging in mental simulation, namely, a concomitant lack of control over one's strategy (Hammond & Summers, 1972). The second result suggests strong personal propensities in evaluating evidence that transcend the particular content of scenarios. Although it is too early to explicate the nature of these individual differences, their existence lends support to the idea that the parameters of our model do capture important aspects of the process that determines judgments under ambiguity.

Fourth, in addition to analyzing output data, we also collected concurrent verbal protocols from several independent subjects who were given selected stimuli from Experiment 1. Subjects were instructed to think aloud as they went through the task, and their responses were tape recorded and transcribed. Although these data can only be considered weak evidence, they are nevertheless revealing about the process we have assumed. For example, in the car-accident scenario, with three witnesses saying the car was green and two saying blue, 1 subject responded as follows:

Five is a pretty small number of people. So, uh, $^3/_5$ths or 60% said it was green. Uh, . . . it seems that it would be, there is a good chance that it could be . . . well, the likelihood that it's green could vary pretty widely. So, I'll say instead of 60%, I'll say 50%.

For the stimulus in which two witnesses said the car was green and one said the car was blue, several responses were of the following type:

Well, first of all there were only 3 witnesses, so that makes everyone suspect. Two said it was green and 1 said it was blue; blue and green are pretty close . . . might depend on what time of day it was, if it were the middle of the day, and if it was bright sunlight, and 2 people said it was green and 1 tends to agree with them; and if it were dusk or night, uh, it seems there would be less certainty that the car was green. So, uh, . . . I would say that it's hard to tell so I'll give it 50%.

I'd put that around 60% sure; there is some agreement but that's only three witnesses, and one of, one third of them didn't agree.

Well, again I would give it a fifty-fifty chance because it's such a small number of witnesses and the colors are similar. So I give it a fifty-fifty chance.

Although subjects are clearly sensitive to the number of witnesses and the similarity of the car colors, they don't always adjust downwards. Consider the following response when 24 witnesses said green and 3 said blue:

Uh, let's see . . . that is $^8/_9$ths which is about 88% said it was green. So I will say given that 88% of the people said it was green, it's very likely that the car was green. 95% likely that the car was green.

Although our model accounts for the rather simple inferences we have studied, it also relates to an important class of inferences that result from surprise. For example, when the credibility of a source is high and the dissimilarity of signals large, one expects the distributions from which p can come to be highly skewed in both directions (recall our earlier example of cameras taking pictures of a bank robber). However, imagine that one camera showed the bank robber to be white and the other showed him to be black. Such a result, where $p = .5$, would be surprising given the credibility of cameras and the dissimilarity of white and black robbers. Indeed, the data are not good enough, given our expectations of the distributions that p could come from. Now consider the low-credibility–low-dissimilarity situation, for example, the taste-test scenario. Imagine that you were told that of the 20 people in the Pepsi versus Coke taste test, all correctly identified the drink as Pepsi. Such a result, where $p = 1$, would be surprising because our expectations of the distributions of p would not contain such extreme values. However, this type of surprise is one in which the data are too good, rather than not good enough. Note that the two types of surprise result when ambiguity is low. Indeed, when ambiguity is high, expectations are weak and surprise (which results from a violation of expectations) is unlikely.

Although our conceptual scheme makes clear when surprise is likely to occur, it cannot handle the variety of possible reactions it can engender. For example, when data are not good enough, it is possible to reduce the credibility of the source (e.g., the cameras were broken), synthesize the hypotheses (there were two bank robbers, one white and the other black), or otherwise make sense of the data by changing the story (e.g., there were two bank robberies on successive days). On the other hand, when data are too good, inferences of fraud, collusion, and the like, are possible (see, e.g., Kamin, 1974, on Burt's twin data; Bishop, Fienberg, & Holland, 1975, on Mendel's pea experiments). An interesting aspect of such inferences is that the surface meaning of the data can suggest the opposite conclusion; for example, consider someone who "protesteth too much," or a suspect who was "framed" for a crime. Indeed, this is implied by our model. Specifically, consider the case of totally unreliable data that imply $\theta = 1$ (see Equation 6a). In this case,

$$S(p_A) = 1 - p_A{}^{\beta}. \qquad (11)$$

Thus, as p_A increases, $S(p_A)$ decreases. More generally, as θ increases, it reaches a point, conditional on p_A and β, where the evidence for a hypothesis starts to be counted against it.

Ambiguity and Risk

Although the importance of ambiguity for understanding risk has been evident since Ellsberg's original article, its omission from the voluminous literature on risk is puzzling. One reason may be the reliance on the explicit lottery, with stated payoffs and probabilities, for representing risky choice. Indeed, as Lopes (1983) noted,

> The simple, static lottery or gamble is as indispensable to research on risk as is the fruitfly to genetics. The reason is obvious; lotteries, like fruitflies, provide a simplified laboratory model of the real world, one that displays its essential characteristics while allowing for the manipulation and control of important experimental variables. (p. 137)

It should be further noted that the explicit lottery has been of equal importance to those interested in axiom systems and formal models of risk.

Although explicit lotteries have been, and continue to be, useful for studying risk, the ambiguities surrounding real-world processes in domains such as nuclear power and environmental safety accentuate the incomplete nature of such representations. Indeed, Ellsberg pointed out the particular importance of ambiguity in understanding people's reactions to new technologies (also see Edwards & von Winterfeldt, 1982, for a historical look at reactions to earlier technological innovations). In any event, the neglect of ambiguity in theories of risk is slowly giving way to interest at both the formal-axiomatic level (e.g., Fishburn, 1983a, 1983b; Gärdenfors & Sahlin, 1982, 1983; Morris, 1983) as well as the psychological level (Ashton, 1984; Lopes, 1983).

Our model of inference under ambiguity has several implications for descriptive models of risky choice. First, because the β parameter can be related to the desirability of outcomes, the model implies that utilities and probabilities are *not* independent. Moreover, Experiment 4 provides direct evidence for this nonindependence. However, the lack of independence only occurs when $\theta > 0$. Thus, whereas the bilinear assumption may be appropriate for models that exclude ambiguity, it is not clear that this assumption can be maintained when ambiguity exists. Second, both our model and data show that the net effect of the adjustment process (i.e., k) varies in magnitude with the level of p_A. Indeed, the net effect of the adjustment can be positive or negative, thereby resulting in regions where $S(p_A)$ is both larger and smaller than p_A. Thus, theories of inference that weight probabilities according to some reliability factor (e.g., Gärdenfors & Sahlin, 1982) cannot account for cross-over effects. Third, the model highlights the difficulty of inferring underlying attitudes toward risk from choices made in ambiguous circumstances. For example, a person buying insurance against a potential loss that is contingent on a small, ambiguous probability might appear risk-averse; however, the same person could appear to be risk-seeking if the probability were larger (cf. Experiment 4). Viewed from the framework of expected utility theory, such behavior would imply an inconsistent utility function. However, this need not be the case because the apparent changes in risk attitude could result from the effects of ambiguity. At the very least, our model provides a way of analyzing the sources of such seemingly inconsistent behavior. As Hogarth and Kunreuther (1984) pointed out, scholars have often attempted to resolve anomalous choice patterns by considering different forms of utility functions. On the other hand, transformations of probabilities have received far less formal attention (for exceptions, see Bernard, 1974; Karmarkar, 1978). Finally, whereas our model does not explicate all aspects of ambiguous choice, it does suggest exciting possibilities for further work in this area.

Extensions to Multiple Sources and Time Periods

To examine inferences under ambiguity in depth, we have restricted ourselves to how evidence from a single source is evaluated at one point in time. However, consider the more realistic situation where decision makers receive information from multiple source types (including base rates) over multiple time periods. The aggregation of information over source types and time can be conceptualized by an *evidence matrix* that has source types for rows and time periods for columns. If we consider J source types ($j = 1, 2, \ldots, J$), and K time periods ($k = 1, 2, \ldots, K$), the entries in the matrix can be denoted by (f_{jk}:c_{jk}), where f_{jk} stands for the number of reports in favor of the hypothesis from source type j in the kth time period, and c_{jk} represents the number of reports against the hypothesis. The evidence matrix provides a simple yet powerful way to look at a wide variety of inference problems. In particular, by focusing on source types (rows) or time periods (columns), one can look at the combining of information either longitudinally, cross sectionally, or both. Furthermore, the issues of reliability and ambiguity become quite complex here inasmuch as there can be differential source reliability, varying numbers of reports per source, and the sources may not be independent. Although the challenge of understanding how people incorporate such factors into their judgments is formidable, the complexity of inferences in real-world settings requires that attention be paid to these issues.

Conclusion

In considering the role of ambiguity and uncertainty in inferential judgments, we have developed a quantitative model that accounts for much existing data as well as our own experimental findings. Furthermore, we have shown how this model relates to Keynes's idea of the weight of evidence, the nonadditivity of complementary probabilities, risky choice, and current work on cognitive heuristics. Moreover, because inference involves "going beyond the information given" (Bruner, 1957), an important way to do this is to construct, via imagination, what might have been or what might be. Such constructions, whether the result of a cognitive-simulation process, as proposed here, or more elaborate processes (as in resolving surprise), pose an interesting and important trade-off for the organism. On the one hand, there are costs of investing in imagination, increased mental effort and the discomfort that results from greater uncertainty. On the other hand, the benefits of considering the world as it isn't, protects one from overconfidence and its nonadaptive consequences. Thus, finding the appropriate compromise between what is and what might have been (or what might be) is central to inferences under ambiguity and uncertainty.

References

Ashton, A. H. (1984). *Ellsberg revisited: The distinction between risk and ambiguity in the real world.* Unpublished manuscript, New York University, Graduate School of Business.

Becker, S. W., & Brownson, F. O. (1964). What price ambiguity? Or the role of ambiguity in decision making. *Journal of Political Economy, 72,* 62–73.

Bernard, G. (1974). On utility functions. *Theory and Decision, 5,* 205–242.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85–94.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Bruner, J. S. (1957). Going beyond the information given. In H. E. Gruber, K. R. Hammond, & R. Jessor (Eds.), *Contemporary approaches to cognition* (pp. 41–69). Cambridge, MA: Harvard University Press.

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

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317–370.

de Finetti, B. (1977). Probabilities of probabilities: A real problem or a misunderstanding? In A. Aykac & C. Brumat (Eds.), *New directions in the application of Bayesian methods* (pp. 1–10). Amsterdam: North-Holland.

Dixon, W. J. (Ed.). (1983). *BMDP statistical software.* Berkeley: University of California Press.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51,* 380–417.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

Edwards, W., & von Winterfeldt, D. (1982). *Public disputes about risky technologies: Stakeholders and arenas.* Unpublished manuscript, University of Southern California, Social Science Research Institute.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32,* 53–88.

Einhorn, H. J., & Hogarth, R. M. (in press). Judging probable cause. *Psychological Bulletin.*

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

Ellsberg, D. (1963). Reply. *Quarterly Journal of Economics, 77,* 336–342.

Fellner, W. (1961). Distortion of subjective probabilities as a reaction to uncertainty. *Quarterly Journal of Economics, 75,* 670–692.

Fishburn, P. C. (1983a). Ellsberg revisited: A new look at comparative probability. *Annals of Statistics, 11,* 1047–1059.

Fishburn, P. C. (1983b, June) Normative theories of decision making under risk and under uncertainty. Paper presented at the Harvard University conference, *Decision Making: Descriptive, Normative, and Prescriptive Interactions,* Boston.

Gärdenfors, P., & Sahlin, N.-E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese, 53,* 361–386.

Gärdenfors, P., & Sahlin, N.-E. (1983). Decision making with unreliable probabilities. *British Journal of Mathematical and Statistical Psychology, 36,* 240–251.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71,* 438–456.

Hammond, K. R., & Summers, D. A. (1972). Cognitive control. *Psychological Review, 79,* 58–67.

Hershey, J. C., Kunreuther, H. C., & Schoemaker, P. J. H. (1982). Sources of bias in assessment procedures for utility functions. *Management Science, 28,* 936–954.

Hogarth, R. M., & Einhorn, H. J. (1985). *Venture theory: A model of risky decision making.* Manuscript in preparation, University of Chicago, Graduate School of Business, Center for Decision Research.

Hogarth, R. M., & Kunreuther, H. C. (1984, October). *Risk, ambiguity, and insurance.* Unpublished manuscript, University of Chicago, Graduate School of Business, Center for Decision Research.

Jepson, C., Krantz, D. H., & Nisbett, R. E. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences, 6,* 494–501.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kamin, L. J. (1974). *The science and politics of IQ.* Potomac, MD: Erlbaum.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance, 21,* 61–72.

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

Koopman, B. O. (1940). The axioms and algebra of intuitive probability. *Annals of Mathematics, 41,* 269–292.

Kyburg, H. E., Jr. (1983). Rational belief. *Behavioral and Brain Sciences, 6,* 231–273.

Levi, I. (1980). *The enterprise of knowledge.* Cambridge, MA: MIT Press.

Lopes, L. L. (1981). *Averaging rules and adjustment processes: The role of averaging in inference* (Report 13). Madison: University of Wisconsin, Wisconsin Human Information Processing Program.

Lopes, L. L. (1983). Some thoughts on the psychological concept of risk. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 137–144.

Marschak, J. (1975). Personal probabilities of probabilities. *Theory and Decision, 6,* 121–153.

Morris, P. (1983, April). *Utility, probability, and decision theory.* Paper presented at the meeting of the Operations Research Society of America/The Institute of Management Science, Chicago.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29–46.

Raiffa, H. A. (1961). Risk, ambiguity, and the Savage axioms: Comment. *Quarterly Journal of Economics, 75,* 690–694.

Rappoport, A., & Wallsten, T. S. (1972). Individual decision behavior. *Annual Review of Psychology, 23,* 131–176.

Roberts, H. V. (1963). Risk, ambiguity, and the Savage axioms: Comment. *Quarterly Journal of Economics, 77,* 327–336.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Schum, D. A. (1980). Current developments in research on cascaded inference processes. In T. S. Wallsten (Ed.), *Cognitive processes and decision behavior* (pp. 179–210). Hillsdale, NJ: Erlbaum.

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

Schum, D. A., & Kelley, C. W. (1973). A problem in cascaded inference: Determining the inferential impact of confirming and conflicting reports from several unreliable sources. *Organizational Behavior and Human Performance, 10,* 404–423.

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

Shafer, G. A. (1976). *A mathematical theory of evidence.* Princeton, NJ: Princeton University Press.

Shafer, G. A. (1978). Two theories of probability. *Philosophy of Science Association, 2,* 1–25.

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

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

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1,* 39–60.

Tversky, A., & Kahneman, D. (1982). Variants of uncertainty. *Cognition, 11,* 143–157.

Wallsten, T. S., Forsyth, B. H., & Budescu, D. (1983). Stability and coherence of health experts' upper and lower subjective probabilities about dose–response functions. *Organizational Behavior and Human Performance, 31,* 277–302.

Winer, B. J. (1963). *Statistical principles in experimental design.* New York: McGraw Hill.

Yates, J. F., & Zukowski, L. G. (1976). Characterization of ambiguity in decision making. *Behavioral Science, 21,* 19–25.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems, 1,* 3–28.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57,* 348–368.

Appendix

This appendix considers (a) the effects of different assumptions concerning the weights given in imagination to values of p greater and smaller than p_A, and (b) an alternative model of the process.

In Equation 4, differential weighting is achieved by the β parameter; that is, $k_g = \theta(1 - p_A)$ and $k_s = \theta p_A{}^{\beta}$. However, one could also consider the class of models where $k(= k_g - k_s)$ is a linear combination of θp_A and $\theta(1 - p_A)$. That is, let

$$k_g - k_s = \theta w_1(1 - p_A) - \theta w_2 p_A$$
$$= \theta[w_1(1 - p_A) - w_2 p_A], \qquad \text{(A1)}$$

where w_1 is the weight given to greater values, and w_2 is the weight given to lesser values. Substituting Equation A1 into Equation 6, we obtain

$$S_l(p_A) = p_A + \theta[w_1(1 - p_A) - w_2 p_A], \qquad \text{(A2)}$$

where $S_l(p_A)$ is used to denote models where k is a linear combination of $\theta(1 - p)$ and θp. Under what conditions does this model imply additivity of the probabilities of complementary events? This is given by

$$S_l(p_A) + S_l(1 - p_A) = p_A + \theta[w_1(1 - p_A) - w_2 p_A]$$
$$+ (1 - p_A) + \theta[w_1 p_A - w_2(1 - p_A)]$$
$$= 1 + \theta(w_1 - w_2). \qquad \text{(A3)}$$

Thus, if $\theta \neq 0$, this class of models implies superadditivity if $w_1 > w_2$, subadditivity if $w_1 < w_2$, and additivity if $w_1 = w_2$. However, this class of models has the property that the amount of nonadditivity is a constant that does not depend on the level of p_A. On the other hand, in the $S(p_A)$ model, the level of p_A does affect the amount of additivity. This is shown in Equation 7, which is reproduced here for convenience,

$$S(p_A) + S(1 - p_A)$$
$$= 1 + \theta[1 - p_A{}^{\beta} - (1 - p_A)^{\beta}]. \qquad \text{(A4)}$$

An alternative model of the process could be given by

$$S_b(p_A) = p_A + \theta(\pi - p_A), \qquad \text{(A5)}$$

where π = person's prior opinion. This suggests a Bayesian-like model where one's current opinion (π) is updated by data. However, this model always leads to the additivity of complementary probabilities because

$$S_b(1 - p_A) = 1 - p_A + \theta[(1 - \pi) - (1 - p_A)], \qquad \text{(A6)}$$

and

$$S_b(p_A) + S_b(1 - p_A) = 1.$$

Received June 15, 1984
Revision received January 15, 1985 ■

Correction to Eich

In the article "Levels of Processing, Encoding Specificity, Elaboration, and CHARM" by Janet Metcalfe Eich (*Psychological Review,* 1985, Vol. 92, No. 1, 1–38), Equation 5 on page 11 should read as follows:

We may summarize the result of retrieval **R** as

$$\mathbf{R} = \mathbf{X}\#(\mathbf{T}) = \mathbf{X}\#[(\mathbf{A}*\mathbf{B}) + (\mathbf{C}*\mathbf{D}) + \cdots]$$
$$= S_{XA}\mathbf{B} + S_{XB}\mathbf{A} + \text{noise}_{A*B} + S_{XC}\mathbf{D} + S_{XD}\mathbf{C} + \text{noise}_{C*D} + \cdots,$$

where **T** is the trace, **A, B, C, D,** and so forth, are list items, S_{XA} is the similarity between **X** and **A**, and **X** is any cue.