DOCUMENT RESUME

ED 287 703 SE 048 658

AUTHOR	Konold, Clifford
TITLE	Informal Conceptions of Probability. Revised.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	Jul 87
GRANT	SED-8016567; SED-8113323
NOTE	48p.; Paper based on a Ph.D. Thesis, University of Massachusetts.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Cognitive Processes; *Cognitive Structures; *College Science; *Heuristics; Higher Education; Learning Processes; Learning Theories; Models; Physical Sciences; Prediction; *Probability; *Problem Solving; Science and Society; Science Education; *Science Instruction
IDENTIFIERS	*Science Education Research

ABSTRACT

This paper illustrates a model of the layperson's reasoning patterns under conditions of uncertainty, the "outcome approach," which was developed from analysis of videotaped problem-solving interviews with 16 undergraduate students. According to the outcome approach, the goal in questions of uncertainty is to predict the outcome of an individual trial. Predictions take the form of yes/no decisions of whether an outcome will occur on a particular trial. These predictions are then evaluated as having been either "right" or "wrong." Additionally, predictions are often based on a deterministic model of the situation. In follow-up interviews using a different set of problems, responses of outcome-oriented subjects were predicted. In one case, subjects' responses were at variance with the "representativeness heuristic." While the outcome approach is inconsistent with formal theories of probability, its components tend to be logically consistent and reasonable in the context of everyday decision making. (Author/TW)

Informal Conceptions of Probability

Clifford Konold

Cognitive Processes Research Group

University of Massachusetts, Amherst

July, 1987

This paper is based on a Ph.D. thesis submitted to the University of Massachusetts. The research was supported by grants SED-8113323 and SED-8016567 from the National Science Foundation.

I am grateful for the continuous support and reactions of Alexander Pollatsek and Amy Robinson and for the helpful comments on earlier drafts by Arnold Well, George Cobb and Harry Schumer.

Requests for reprints should be sent to Clifford Konold, Cognitive Processes Research Group, Hasbrouck Laboratory, University of Massachusetts, Amherst, MA 01003.

Abstract

A model of the layperson's reasoning under conditions of uncertainty, the <u>outcome approach</u>, was developed from analysis of videotaped problem-solving interviews with 16 undergraduates. According to the outcome approach, the goal in questions of uncertainty is to predict the outcome of an individual trial. Predictions take the form of yes/no decisions of whether an outcome will occur on a particular trial. These predictions are then evaluated as having been either "right" or "wrong." Additionally, predictions are often based on a deterministic model of the situation. In follow-up interviews using a different set of problems, responses of outcome-oriented subjects were predicted. In one case, subjects' responses were at variance with the "representativeness heuristic" (Kahneman & Tversky, 1972). While the outcome approach is inconsistent with formal theories of probability, it's components are logically consistent and reasonable in the context of everyday decision-making.

Informal Conceptions of Probability

A weather forecast includes the prediction of 70% chance of rain; lotteries and sweepstakes publish the odds of winning; Consumer Reports publishes frequencies of types of repairs for various models of cars. Information of this sort is intended to help people make more reasonable decisions. Yet, recent research on human decision making in situations involving uncertainty has revealed that peoples' judgments are frequently not in agreement with probability and statistical theory (Kahneman, Slovic & Tversky, 1982; Peterson & Beach, 1967; Pollatsek, Konold, Well, & Lima, 1984.)

Amos Tversky and Daniel Kahneman have provided the most integrative account to date of the discrepancies between normative and actual judgments under uncertainty. According to Tversky and Kahneman (1983), two general types of cognitions are potentially available in making probabilistic judgments. On the one hand, people have acquired some knowledge of random events and basic probability theory that allow them to calculate the chance of various events in simple chance setups. Most people know, for example, that $p(A) + p(A) = 1$ and that for setups with equally likely outcomes the probability of a particular event is equal to the number of outcomes favorable to that event divided by the total number of equally-likely outcomes (the classical interpretation of probability). Piaget and Inhelder (1951/1975) concluded that by the age of 12 most children can reason probabilistically about a variety of random generating devices.

In addition to these capabilities, however, people have developed a number of judgment heuristics for analyzing complex, real-world events. These heuristics, according to Tversky and Kahneman (1983) are based on a collection of "natural assessments" that operate on a non-conscious, perceptual level.

While most decisions based on natural assessments are congruent with those that would be made on the basis of probability theory, there are many situations for which this is not the case, when the perceptual processes and associated judgment heuristics lead to "statistical illusions."

For example, most people incorrectly believe that the sequence MMMMMM of male and female births in a family is less likely than the sequence MFFMMF. Kahneman and Tversky (1972) suggest that this conclusion is reached through application of the "respresentativeness heuristic," according to which the probability of a sample is estimated by noting the degree of similarity between the sample and parent population. Since the sequence MFFMMF is more similar to the population proportion of approximately half males and half females and also better reflects the random process underlying sex determination, it is judged as more likely.

Asked to compare the frequency of words in the English language that begin with r to those that have r as the third letter, people typically conclude, and incorrectly, that the former are more frequent. According to Tversky and Kahneman (1973) this judgment is made via the "availability heuristic" according to which the probability or frequency of an event is related to the ease or difficulty of recalling relevant instances of that event. Since it is easier for most people to mentally search for words according to their first letter, they mistakenly judge them as occuring more frequently.

When making probabilistic decisions both the collection of natural assessments and more formal, conceptual knowledge of probability theory are presumably available. Which of these is applied in a particular instance is a function not only of individual differences in knowledge of probability theory, but also of situation variables. Nisbett, Krantz, Jepson, and Kunda

(1983) have shown that people with little formal training in probability will tend to analyze a situation probabilistically when (a) the sample space is easily recognizable, as when the event is repeatable and outcomes are symmetric, and (b) the role of chance is salient, as in coin flipping and urn drawing. On the other hand, even people who have had considerable training in the application of probabilitistic models can be led to the unconscious application of natural assessments for situation which they know call for a probabilistic analysis (Tversky & Kahneman, 1971).

Formal and Informal Conceptions of Probability

In this paper I explore the possibility that errors in reasoning under uncertainty arise not only from indiscriminate application of natural assessments, but also from analyses based on conceptual knowledge that is inconsistent with formal probability theory. Evidence for these types of conceptual errors was sought by examining subjects' verbalizations as they reasoned about various situation involving uncertainty. On the basis of subject statements, a model of reasoning under uncertainty was formulated. According to this model, referred to as the outcome approach, the goal in dealing with uncertainty is to predict the outcome of a single, next trial. For example, subjects given an irregularly shaped bone to roll and asked which side was most likely to land upright, interpreted the question as a request to predict the outcome of a single trial. Subjects' evaluated their predictions as being correct or incorrect after the results of the single trial. Furthermore, predictions in the outcome approach are often based on a causal analysis. Numbers that are assigned as "probabilities" may gauge the strength of these causal factors, but more typically are used as modifiers of the yes/no prediction, with 50% meaning that no sensible prediction can be made.

The outcome approach differs from formal theories of probability and will be contrasted in particular to the frequentist and personalist interpretations. To the frequentist, a probability is meaningful only with respect to some repeatable event and is defined as the relative frequency of occurrence of an event in an infinite (or very large) number of trials (Reichenbach, 1949; von Mises, 1957). This is viewed as an objective theory in that the probability is regarded by the frequentist to refer to an empirical, verifiable quantity. A rival subjective theory is the personalist interpretation (deFinetti, 1972; Savage, 1954) which holds that a statement of probability of some event communicates the degree of belief of the speaker (measured by the amount that would define a "fair bet") that the event will occur.

Though theorists quibble amongst themselves over whether some event ought to be assigned a probability, and over the interpretation of the probability, the various schools generally derive identical probabilities for events they all agree are probabilitistic. For example, the probability in coin flipping of the outcome heads would be determined as .5 on the basis of the classical interpretation since the ratio of favorable to total number of equally-likely alternatives is 1 to 2. For the frequentist it would be .5 if the limit of the relative frequency of heads approaches .5 as the number of trials approaches infinity. Presumably, this would occur given a fair coin. According to the personalist interpretation, different people could validly assign different values to the probability of a particular coin based on their beliefs about the fairness of the coin, the character of the person doing the flipping, the technique of flipping, etc. However, in formalizing a personalist view theorists have included various adjustment mechanisms that require the revision of initial probabilities given new information about the actual occurrence of the event. Savage (1984), for example, advocates the use of Bayes' Theorem to revise initial beliefs. Given enough data about the frequency of occurrence of heads of a particular coin, subjective probabilities are thus constrained to converge on the frequentists' limit. It is at this level that the outcome approach will be contrasted to formal theories of probability. That is, the outcome-oriented individual does not regard frequency information as relevant in cases where formal theories would all agree that it is.

Overview of Study

In this study subjects were interviewed on two occasions. In Interview 1, a set of questions dealing with various aspects of probability were given to 16 subjects. Videotapes of these interviews were analyzed, and aspects of subjects' reasoning that were at variance with formal probability theory were identified. Proceeding on the assumption that there were logical connections between various statements that subjects made, a two-feature model of their reasoning, the outcome approach, was developed. Responses that could be regarded as indicators of reasoning consistent with features of the outcome approach were then coded. On the basis of this coding a score was generated for each subject that reflected the degree of adherence to the outcome approach. Interview 2 was then conducted to test the predictive validity of the outcome approach. The same subjects were given another set of problems for which specific predictions had been made on the basis of their performance in Interview 1. These data are used to support the argument that peoples' beliefs about various aspects of probability while non-normative are interrelated--that there is an internal coherence to their beliefs.

Interview 1

Method

Subjects

Interview 1 was undertaken to identify aspects of subjects' reasoning that were non-normative yet used consistently across a variety of problems involving uncertainty. Sixteen undergraduate students at the University of Massachusetts at Amherst were interviewed as they attempted to solve word problems that involved uncertain outcomes. Subjects volunteered their participation in return for extra course credit in a psychology course.

Problems

The three problems and follow-up questions that were used in Interview 1 are presented below in an abbreviated form. (The problems in their entirety are included in the Appendix.)

> Weather Problem. What does it mean when a weather forecaster says that tomorrow there is a 70% chance of rain? Suppose the forecaster said that there was a 70% chance of rain tomorrow and, in fact, it didn't rain. What would you conclude about the statement that there was a 70% chance of rain? Suppose you wanted to find out how good a particular forecaster's predictions were. You observed what happened on ten days for which a 70% chance of rain had been reported. On three of those ten days there was no rain. What would you conclude about the accuracy of this forecaster?
>
> Misfortune Problem. I know a person to whom all of the following things happened on the same day. First, his son totalled the family car and was seriously injured. Next, he was late for work and nearly got fired. In the afternoon he got food poisoning at a fast-food restaurant. Then in the evening he got word that his father had died. How would you account for all these things happening on the same day?
>
> Bone Problem. I have here a bone that has six surfaces. I've written the letters A through F, one on each surface. If you were to roll that, which side do you think would most likely land upright? How likely is it that x will land upright? (Subject is asked to roll the bone to see what happens.) What do you conclude about your prediction? What do you conclude having rolled the bone once? Would rolling the bone more times help you conclude which side is most likely to land upright?

The problems were selected to vary along several dimensions and can be categorized according to criteria mentioned by Nisbett et al. (1983). The Bone Problem involves a reasonably clear sample space, evident repeatability of trials, easily identified chance factors, and strong cultural prescription toward viewing the phenomena statistically. The Misfortune Problem is low on all these dimensions. The Weather Problem is intermediate in the clarity of sample space and cultural prescription, and low on repeatability of trials and identifiable chance factors.

Procedure

I interviewed subjects individually in a session lasting approximately one hour. Subjects were instructed that they would be g en several problems that would require reasoning about situations involving uncertainty. They were told that the particular answers they gave were of less interest than the reasoning that led to the answer. Accordingly, they were instructed to "think aloud" as they attempted to solve each problem, verbalizing their thoughts as they occurred rather than attempting to reconstruct them at some later time. A felt pen and pad of paper were provided for the subjects' use. Subjects were informed that the interview would be videotaped, and the recording equipment was in full view.

The problems were presented orally. Two orders of presentation were used, the order being alternated on each successive interview. Order A was the sequence Weather, Bone, Misfortune. Order B was th reversed sequence.

The majority of probes used during the interview consisted of requests to repeat a statement and reminders to verbalize. However, unplanned probes were used occasionally in an attempt to further elucidate subject's thinking. The interview format, therefore, could best be characterized as "in depth" (Konold & Well, 1981) as opposed to "think aloud" (Ericsson & Simon, 1984).

Results and Discussion

A qualitative analysis of the interview protocols suggested that a subset of subjects were reasoning according to a non-normative, yet coherent, belief system. This system, the outcome approach, can be characterized as involving two general features:

(a) the tendency to interpret questions about the probability of an outcome as a request to predict the outcome of a single trial;

(b) the reliance on causal as opposed to stochastic explanations of outcome occurrence and variability.

To give an initial impression of the outcome approach, two composite interviews are juxtaposed in Table 1. On the left is a prototype of the outcome approach; on the right, a prototype of a frequency interpretation. These prototypes are assemblages of excerpts from several subjects (as noted) and should be regarded as ideal characterizations. Only a few of the subjects' protocols closely resemble one or the other of these prototypes.

Insert Table 1 about here

In the remainder of this section the two features will be more formally described and exemplified by referring to numbered excerpts in Table 1.

Predictability of Individual Trials

Two types of statements indicated that some subjects perceived their goal as predicting outcomes of individual trails. These statements consisted of (a) qualitative, yes/no predictions and (b) right/wrong evaluations of predictions.

Qualitative predictions. In the outcome approach, predictions of individual trials take the form of "yes," "no," and occasionally "I-don't-know" decisions of whether or not a particular outcome will occur. This contrasts with the frequency interpretation where typically the objective is to predict a global index of the entire sample such as the mean or percent of some outcome in a series of trials. Four of the subjects translated the statement "70% chance of rain" into the more definitive and qualitative statement, "It's going to rain." This translation was usually accomplished by using the range of 0% to 100% as a decision continuum, with 0% meaning "no," 100% as "yes," and 50% as "I don't know." Intermediate values were ultimately associated with one of these three anchor or decision points according to a vague and variable proximity criterion. Thus, 70% was considered significantly above 50% to warrant identification with 100%, or "yes," with perhaps some associated expectation of error (see excerpt 2). Given this qualitative (yes/no) interpretation of the probability range, 50% was not viewed as a predictive forecast by three of the subjects, but as an admission by the forecaster of total ignorance about the outcome. For example, Subject 9 replied:

> S9: It's not 100% chance and it's not 50/50, so he's not guessing. If he said 50/50 chance I'd kind of think that was strange...that he didn't really know what he was talking about, because only 50/50--"it might rain or it might be sunny, I really don't know."

<u>Evaluation of predictions.</u> That subjects see their task as predicting individual trials is further indicated by the tendency to evaluate a prediction as having been either right or wrong after the occurrence of a single trial.

In the Weather Problem a situation was posed in which no rain fell on a day for which a 70% chance of rain had been estimated. Asked what they would conclude about the accuracy of the statement that there was a 70% chance of rain, six of the subjects responded that the statement must have been incorrect (see excerpt 3). Subjects were also questioned about the accuracy of a forecaster who had predicted 70% chance of rain for ten days, when in fact no rain was recorded on three of the ten days. Theoretically, seven days of rain out of ten is the most likely outcome given an accurate 70% forecast on each day. Three of the subjects' responses were consistent with this reasoning. Nine of the subjects concluded, however, that the forecaster was only "pretty accurate," suggesting that there was room for improvement (see excerpt 4). Four subjects expressed a conflict over whether the forecaster was perfectly accurate or not. At the heart of this conflict was the question of whether the forecaster is trying to formulate (a) an accurate prediction of the relative frequency of rainy days, or (b) a decision about whether or not it will, in fact, rain. Subject 8 concluded:

> S8: Well, he's looking at an individual day--particular day--and he's setting up percentages on one day. And you can't really extend that to an amount of time, I don't think.

The tendency to evaluate outcome predictions as either right or wrong was also evident in the Bone Problem. After making an initial guess of which side of the bone was most likely to land upright, subjects were asked to roll the bone. Nine of the subjects remarked that their guess was either right or wrong having observed the result of one trial (see excerpt 9).

Evidence from both the Bone and Weather Problems supports the claim that a subset of subjects encode requests for probabilities as requests for a decision of which alternative will occur on a particular trial. Once the trial has been conducted, these predictions are retrospectively evaluated as having been either right or wrong. When probabilities are provided, as in the Weather Problem, they are not interpreted as probabilities per se, but as values that can be used to formulate a yes/no decision.

Predicting Outcomes from Causes

In this section evidence is presented to support the claim that individuals frequently arrive at or interpret estimates of probability through a causal analysis of the situation. It needs to be stressed that a formal probabilistic approach does not necessitate the denial of underlying causal mechanisms in the case of chance events. Hypothetically, one can imagine describing the last in a series of 100 tosses of a fair coin in sufficient detail such that it could be seen to be determined by events that preceded it. In practice, however, a causal description is often seen as impractical if not impossible (e.g., von Mises, 1957, p. 208-209). Accepting a current state of limited knowledge, a probabilistic approach adopts a "black-box" model according to which underlying casual mechanisms, if not denied, are ignored. The mechanistic model is not abandoned in the outcome approach. The goal of predicting the results of individual trials in a yes/no fashion would, in fact, seem to imply the possibility of determining beforehand the results of each individual trial.

Weather Problem. In the Weather Problem subjects were asked to explain the meaning of the number in the proposition, "There is a 70% chance of rain." Four subjects suggested that the 70% was a measure of the strength of a factor that would produce rain (e.g., 70% humidity or 70% cloud cover; see excerpt 1). Three subjects used causal explanations to account for the non-occurrence of rain given the forecast of 70% chance of rain (see excerpt 3).

Misfortune Problem. Eight subjects gave other than chance explanations of the several low-probability events in the Misfortune Problem. Six subjects tried to embed all of the events in a causal sequence such that each could be seen to have been a direct result of a preceding event (see excerpt 6). Five subjects relied on explanations that involved causal agents such as God or the stars.

Bone Problem. In the Bone Problem, five subjects expressed reservations about whether additional trials would be helpful in determining which side was most likely to land upright (see excerpt 10). Three of these subjects suggested that more reliable information could be obtained from careful inspection of the bone than from conducting trials. Three subjects did not use the data provided from the results of 1000 trials in predicting the results of 10 trials. Eight subjects attributed variations among trials to the way the bone was rolled.

To summarize, a variety of subject statements suggest an informal approach to probability for which predicting the result of an individual trial is the primary goal. Arriving at a prediction often involves an analysis of causal factors. Numeric values that may be associated with a prediction are measures of the confidence that the predicted outcome will occur, as well as measures of the strength of relevant causal factors. When probabilities are given to the outcome-oriented individual, they are recoded into a "yes," "I-don't-know," or "no" decision according to their distance from the corresponding values of 100%, 50%, and 0%. Table 2 is a summary of statements made by subjects that were indicative of the outcome approach.

Insert Table 2 about here

Outcome scores at the bottom of the table were determined for each subject by summing the number of categories checked in Table 2. Scores had a possible range of from 0 to 15, with higher scores being indicative of an outcome orientation. The median for the 16 subjects was 4.17.

Insert Table 2 about here

------------------------- ----

To test the validity of the outcome approach, a second set of interviews was conducted. Specific predictions were made (see below) concerning the responses of the same subjects to a different set of problems.

Interview 2

Method

Subjects

Twelve of the original sixteen subjects returned to participate in the follow-up interviews. The other four could not be located. Approximately five months had elapsed between Interviews 1 and 2.

Problems and Procedure

Four problems were employed. The Cab Problem has been used in previous research (Kahneman & Tversky, 1972). The remaining three problems were developed and then standardized in 14 pilot interviews. All four problems are presented below in abbreviated form and in the order the order they occurred in the interview. The problems are presented in their entirety in the Appendix.

Cab Problem. (Subject is asked to read the Cab Problem aloud.) "A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

(i) 85% of the cabs in the city are Green and 15% are Blue.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs, half of which were Blue and half of which were Green, the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

What is the probability that the cab involved in the accident was Blue rather than Green?"

Bone-2 Problem. Last time you were asked which side of this bone you thought would most likely land upright? Do you remember which side you concluded? (The bone is held far enough away so that the labels cannot be read.) I'm going to ask you the same question again. And to give you something to base your answer on, I'll offer you any one of the following pieces of information. (Subject is shown the list as the interviewer reads the items.)

1 - A measure of surface area of each side.
2 - The results of 100 rolls made by 16 people.
3 - The results I got in 1000 rolls.
4 - A drawing of the bone showing the center of gravity.
5 - The bone to look at.
5 - The results of your last 10 rolls.

Painted-die Problem. I have here a six-sided die. Suppose I painted five of the surfaces black and the other one white. If I rolled the painted die six times, would I be more likely to get six blacks or five blacks and one white? If I rolled it 60 times, how many times would you expect the white surface to come up?

Modeling Problem. Would there be a... way that we could make a model of the bone so that instead of rolling the bone, we could pick something out of a container and get the same kind of results? (If a subject cannot generate a model, four possible models are suggested in succession, and the subject is asked to comment on their appropriateness. When, and if, subjects agree upon a model of the bone, they are asked the following questions:) Suppose I rolled the bone 100 times and kept track of what I got, then I draw 100 times from this can filled with the labeled stones. If I showed you the results from both, could you tell from looking at the results, which I got from rolling the bone and which from drawing from the container? In those 100 trials with the bone and the container, do you think with one of those I'd be more likely than with the other to get no E's? Do you think I'd be more likely with one of those to get more D's in 100 trials than with the other?

Initial instructions to subjects were similar to those given in Interview 1. Subjects were told they would be given several problems that involved uncertain outcomes. They were reminded to "think aloud" and to use the pen and paper for any figuring they might want to do. All the problems except the Cab Problem were presented orally, and the entire interview required approximately 40 minutes.

Results and Discussion

The four problems used in Interview 2 were designed to determine whether the responses of outcome-oriented subjects to another set of questions could be predicted. To test these predictions, scores based on performance in Interview 2 were correlated with the outcome score that summarized subjects' performance in Interview 1. For the 12 subjects who were interviewed on the second occasion, outcome scores ranged from 0 to 13, with a median of 4.5.

While the full rationale for the four problems will be made clear in the subsequent discussion, a brief summary follows:

(a) Cab Problem: In prior research using this problem, subjects had made statements consistent with the decision and single-trial features. The problem was used as an independent measure of the consistency of these features over problems and sessions.

(b) Bone-2 Problem: Subjects were again asked to predict outcomes of rolling the same bone that had been used in Interview 1. A different set of probes was used to determine whether estimates were being generated primarily from frequency information or from physical features of the bone.

(c) Painted-die Problem: Given that the unit of analysis in the outcome approach is the single trial, it was predicted that outcome-oriented subjects would solve the problem by first imagining the results of each individual trial and then summing these results together to obtain the solution for six trials. Since black would be the best guess for each individual trial, it was predicted that outcome-oriented subjects would believe that six blacks is more likely than the normative solution of five blacks, one white.

(d) Modeling Problem: This problem was designed to test the validity of the casual feature. It was predicted that outcome-oriented subjects would not believe that a urn model could be constructed that could duplicate the results of rolling the bone since salient causal features had been altered.

In the remainder of this section, each problem and the associated predictions will be discussed in turn. After specifying the predictions that were made prior to conducting the interviews, correlations between performance in Interviews 1 and 2 will be reported, and then selected excerpts from the interviews that pertain to the predictions will be presented.

Cab Problem

The Cab Problem (originally used by Kahneman and Tversky, 1972) has been used to study subjects' reluctance to take into account base rates (in this case the relative number of the two colors of cabs) in the formulation of probability estimates. Well, Pollatsek and Konold (1983) using an interview format reported that many subjects believed that they were not being asked the probability that the errant cab was blue, but whether or not, in fact, it was blue. In addition, numeric answers that subjects were asked to provide in many cases seemed to be only loosely based on the numbers given in the problem. These observations are similar to subjects' statements in the Bone and Weather Problems from which the decision and single-trial features of the outcome approach were inferred.

Given that the outcome approach describes a general orientation to uncertainty, those subjects who responded in an outcome-oriented fashion in Interview 1 should respond in a similar way to the Cab problem. Specifically, it was predicted that outcome-oriented subjects, as defined by higher outcome scores in Interview 1, would be more likely to:

(a) ask whether a number was required in answering the question of the probability that it was a blue cab;

(b) encode the question, "What is the probability...?" into the question, "What color was the cab?" (this encoding being indicated by responses such as, "I think the cab was blue"); and

(c) base a numeric answer on a "loose" or qualitative interpretation of the evidence they thought relevant.

Individual scores for the Cab Problem could range from 0 to 3 and were obtained by summing the number of predicted responses for each subject as indicated in Table 3. Coders for this (as well as the other three) problems were myself and a graduate student who was blind both to the nature of Interview 1 and to the hypotheses being tested. Interrater reliability for coding the three categories for the Cab Problem was estimated by correlating the set of ratings of the two coders, with $\underline{r} = .759$. The scoring rule applied was that both coders had to agree that a particular statement had been made in order for it to be counted. These scores were correlated with the outcome scores from Interview 1 with $\underline{r} = .586$, $p < .025$ (one-tailed).

Insert Table 3 about here

Given that the goal of the outcome approach is to determine what will or did occur, the question concerning probability is translated into the question, "What happened?" as indicated in the following response:

> S1: So you want to know if I think that's right -- if it was blue. Well, I would say it would be blue rather than green -- just the fact that this really isn't important -- the 85% are green, 15% are blue. I mean there are still a substantial amount of blue cabs out there. But the fact that the guy said -- well, the court said that "in 80% of the cases you identified the right color." And the guy said he saw blue. He doesn't say "I think I saw blue." He says, "I saw blue." So I would go with blue.

In the Cab Problem, subjects were asked specifically for the probability that the cab was blue. A subject's query of whether a number was required was considered consistent with the decision feature of the outcome approach. When subjects asked if a number was wanted, I hesitated in order to allow them to clarity the question, and then if they did not continue, I asked what the alternative was to giving an answer:

> S11: Let's see. Am I looking for a number as opposed to like -- Am I looking to say, "It's 80% probability that it was a blue rather than green?" Is that what I'm --
>
> I: What's the other option? How else would you prefer to give that?
>
> S11: Sure, it could have been a blue cab. (Laughter) No -- just that it would have been a strong -- it was more likely as opposed to less likely. Kind of like this fit in. More positive as opposed to a definite number positive.

Central to the goal of specifying what will happen or did happen is the focus on single trials: Questions of uncertainty are viewed as pertaining to a particular event as opposed to a set of events. Subject 5 justified ignoring the base-rate information on the grounds that at issue was the occurrence of a particular event, and that information regarding a class of events was irrelevant:

> S5: It really doesn't matter how many cabs there are in the city. What you're thinking about is this one particular cab, whether it was blue or green. And since the guy was usually right, he's probably right.

As suggested in the above excerpt, the witness identification can be seen as applying to the individual event (the color of the errant cab) in a way that the base-rate information cannot. Using the base rates would seem to require regarding the particular accident as one of a set of accidents involving the two cab companies. To the outcome-oriented individual, this is not relevant to the question; what matters is this particular accident. It is evident in the above and following excerpts that the witness identification is not viewed as one of a class of similar identifications. Rather, the outcome-oriented individual may assign the attribute "pretty reliable" to the witness and thus to the witness's identification of the color of the errant cab based on the accuracy data collected by the court. It may be in the process of assigning this attribute that subjects "let go" of the specific meaning of the 80% and then give a confidence value for their belief that the cab was blue which is only loosely based on the 80% estimate of the witness's accuracy.

S8: And since his visibility was pretty clear, and just on that -- I'm not even taking these numbers so much as just, you know. conceptualizing it. Since he saw it was blue, and there's more of a chance that he's right as seeing it as blue, that he saw it correctly. So I'll say that.

S3: 80% just because he had -- his percentage correct before was 80%, so it makes sense that he, probably -- chance 80% that he got it right this time.

I: OK.

S3: Maybe better.

I: Can you explain why you think it might be better than that?

S3: Well, because more than not he got them right when they tested him before. So that's why it would be possible that he'd be more than 80%.

S13: Yeah -- that he did guess, more than he didn't, the right colors. So I'd go with the blue. I'd say that it was a blue one.

I: And how about just an estimate of what the probability would be, or a guess.

S13: I want to say just 80...

I: Is that 80 based on this [pointing to 80% witness accuracy]?

S13: No. I'm just trying to find -- I'm just trying to think of something that's closer to 100 -- like over to more of a chance that it happened.

<u>Bone-2 Problem</u>

Given another opportunity to decide which side of the bone was most likely to land upright, it was predicted that outcome-oriented subjects would prefer to consider the physical features of the bone rather than frequency data. It was also predicted that, when asked how a statistician would determine the probabilities associated with each side of the bone, outcome-oriented subjects would believe that various physical features of the bone would be taken into account.

Scores for this problem could range from 0 to 4 based on the following four categories of subject responses:

(a) the first choice for analyzing probabilities associated with the bone was not frequency data;

(b) the second choice was not frequency data;

(c) physical properties of the bone were used in predicting the results of ten trials;

(d) It was believed that a statistician would consider physical properties in determining probabilities associated with rolling the bone.

The interrater reliability for coding statements as indicative of these categories was 100%. The correlation between scores on the Bone-2 problem and the outcome scores was $\underline{r} = .782$, $p < .005$ (one-tailed).

As predicted, outcome-oriented subjects were more likely to believe that a decision about the probability of various sides of the bone to land upright should be arrived at by considering the physical features of the bone. One hypothesis as to why these subjects prefer a physical to a statistical analysis is that the physical features of the bone might be viewed as a more stable source of evidence when compared to frequency data which can fluctuate from sample to sample. This seemed to be the rationale given by Subject 12 for basing predictions on an inspection of the bone. Asked why she thought the data from 100 rolls were unimportant, she replied,

> S12: Well, because what they did may not be -- it's sort of chance, you know, that happened. If the same 16 people did the same 100 rolls, it would probably be different the second time. It just doesn't seem a very specific kind of statistic.
>
> I: And why do you think it would be different?
>
> S12: Things change. I don't think anything duplicates itself exactly the second time.
>
> I: How about the results I got in 1000 rolls?
>
> S12: Yeah, that too is kind of iffy. If you did the same thing over again, plus a second 1000 rolls -- I mean, you could go on for 2000 rolls or whatever, and I don't know if it really would tell you much. Then again, I could be wrong.

A second hypothesis as to why a physical analysis might be preferred is that physical properties may be viewed as casual agents of what one wants to predict, while frequency information is not. However, the interviews provided no compelling evidence that this is the case. One subject did express the belief in Interview 2 that the physical properties were "real evidence" in contrast to frequency data. Asked to explain how she decided that D was the most likely side, she responded,

> S3: Well, just 'cause it's flatter on the underside, so it's more likely to land on that side than it would on any other place.
>
> I: Are you using this information at all [the results on her last ten rolls]?
>
> S3: Maybe a little, yeah. I suppose. Well, I looked first and thought that was reasonable. So--

Asked how a statistician would determine the probability, she first mentioned surface area. Asked if they would use anything else, she replied, S3: Well, they would probably make rolls themselves and see how it comes up. But I don't know if they would use that for real evidence or whatever.

I: You feel like the results of what you got isn't real evidence?

S3: Well, yeah. It has some. But there must be some, you know, like measuring the side, and that must be a little more precise than my rolls.

This subject's last statement suggests that she regards the properties of the bone as a more valid type of evidence because they are easier for her to think of as being measured precisely.

Subjects 5, 11, and 13, who also considered features of the bone to be important in a determination of probabilities, suggested that they should be used in conjunction with, rather than to the exclusion of, frequency data:

S5: I'd take number three [n = 1000 data], and I'd look at each surface of the bone that had come up and compare it to the number of times it had gotten up and see why it had so I could decide whether or not the results were accurate, according to the shape of the bone.

In predicting ten rolls, Subject 5 inspected the bone carefully to decide how he would allot predicted frequencies to B and C since, according to him, they were so close in frequency of past occurrence. His explanation of how a statistician would estimate probabilities was consistent with the approach he had employed:

S5: A statistician would count a great deal of weight to the center of gravity and how it related, and, taking your results [n = 1000 data], would come up with a bunch of statistics that would probably reflect fairly accurately your results, with perhaps some modification according to what he thought the structure of the bone gave out.

Subject 11 used only frequency data to make predictions about the bone, but expressed the belief that a statistician would, in a "joint effort," supplement these with an analysis of physical properties:

> S11: 'Cause you'd roll the bone and get a rough idea of the probabilities, whatever they are -- yeah, probabilities -- and take it to have it analyzed to figure out if, structurally, you can understand why these-- you know. You assign these particular values to each face, and then through comparing both, just --
>
> I: But I might want to modify what I had got rolling it?
>
> S11: Yeah. It's just kind of like added significance, or not significance -- added sureness, or whatever, -- belief in your percentages.

In summary, the tendency to view physical properties of the bone as important in the determination of probabilities of the various landing orientations is strongly related to measures of the outcome approach. Physical properties appear to be regarded as information at least on a par with frequency data in making predictions.

The correlations between performance on the first interview and the first two problems of Interview 2 suggest that subjects' outcome-oriented responses are consistent over time. The last two problems involved using the outcome approach to anticipate specific responses that had not been observed in Interview 1. Thus, they provide more compelling evidence of the validity of the outcome approach.

Painted-die Problem

In the Painted-die Problem, subjects were first presented a die and then six stones, both of which consisted of five elementary outcomes of one type (black) and one of another (white). They were asked to predict whether in six trials they would be more likely to observe five blacks and one white, or six blacks. Theoretically, the former is more likely, the probability of exactly five blacks being .402, the probability of six blacks being .335.

Most people, when asked, will respond that the probability of white being rolled is one out of six. But it is not clear what is meant by such an answer other than that there is only one white out of six sides. If people viewed "one out of six" consistent with formal probability theory they would expect to get, on average, one white in six trials, which is also the model outcome. Even failing this line of reasoning, one would predict on the basis of the "representativeness heuristic" that people would believe five blacks to be the more likely outcome since it looks more like, and in this case, is identical to, the population distribution. Kahneman and Tversky (1973) reported results on a similar problem involving drawing cards with replacement from a deck in which 5/6 of the cards were marked X and the remaining 1/6 were marked O. They indeed found that subjects judged five X's and one O to be more likely than six X's.

It was predicted that outcome-oriented subjects, however, would regard six blacks as the more likely outcome. In the outcome approach, the primary unit of analvsis is the individual trial. Application of the representativness heuristic in this problem requires a focus on predicting the sample result rather than the individual trial results. Given a probability value, the outcome-oriented individual arrives at a prediction of a trial by deciding which yes/no or I-don't-know decision point the probability value is closest to. Thus, rather than viewing the 5/6 as a value that relates to the expected relative frequency of blacks in randomly drawn samples, it was predicted that outcome-oriented individuals would give it a qualitative interpretation of the approximate form "the next trial will almost certainly result in a black." When asked to predict the outcome for six trials, rather than using the 5/6 to form an expectation for the set of six trials, they may arrive at a prediction by summing over their expectations for each of the six trials. Since this expectation is more qualitative than quantitative in nature, it was expected that outcome-oriented subjects would more frequently say that six blacks are more likely, and that they would also believe that the ratio of blacks to white over a larger series of trials will remain above the normative value of five to one.

Scores for the Painted-die Problem had a range of 0 to 3 based on the following three categories:

(a) six blacks stones were judged as more likely than 5 black, one white;

(b) fewer than 10 white were expected in 60 trials, or, on the average, more than six trials were required to roll one white;

(c) the probability of a black on the first trial was estimated to be above 5/6 or above 84%.

Interrater reliability for coding the Painted-die Problem was 100%. The correlation between scores on this problem and outcome scores from Interview 1 is $\underline{r} = .616$, $p < .025$ (one-tailed).

Excerpts from the interviews indicate that, as suggested, subjects solved the problem by imagining a single trial for which the probability of black is overwhelming, and then extended this prediction over trials to arrive at the conclusion that six blacks was the more likely outcome.

Subject 7 initially stated that the probability was 5/6 for a black. Later, however, he stated that six blacks were more likely and that ten or fewer whites would occur in 60 trials:

> S7: Well, I think it's -- the white's there, but -- I'm not exactly sure what I'm trying to say. Just because the odds are always the same. There's only one of them in there. So even though

> it's six rolls and there's six things in there. there's only one or the other that's going to come up each time. And that -- chances are better than five to one, one of the five blacks is going to come up.

Similar reasoning is demonstrated by Subject 15:

> S15: Because it's a higher probability of getting a black side because there are more black sides and so there's more probability that when you roll it, you're going to get a black side instead of that one white side.

Subject 3 combined the "more blacks" rationale with the reasoning that the sampling with replacement procedure does not guarantee white:

> S3: Probably more likely to get all black just 'cause -- I don't know what percentage, but most of the die is black, so it's going to come up on that side. 'Cause you're not going to roll it on a different side each time you roll it, so that it's bound to come up one of those six rolls. So it probably would be black on all of them.

Subject 5 believed that rolling six dice at once would result in five blacks, but that rolling the same die six times would result in six blacks:

> S5: Well, each roll is a separate entity. You roll it, and a side will come out. You don't roll all six at one time. So likelihood is that each time it comes out, the side that has the dominate color, which is black, is the color that'll come out.

He finally rejected this reasoning, favoring five blacks in both cases. His initial response, however, provides a good example of what is being regarded as the outcome approach to this problem -- that of imagining the results of one trial as almost certainly being black, and, by extending this qualitative judgment, concluding that six blacks are more likely over six trials. It is especially significant that this subject began thinking differently about the problem when he imagined all six trials occurring at once, changing his focus from six, single trials to a set of trials. (A similar belief in a difference between flipping one coin repeatedly and several at once was defended by the 18th century mathematician, D'Alembert. For an interesting account of this and other of D'Alembert's unconventional beliefs about probability, see Todhunter, 1949.)

Modeling Problem

The modeling problem was designed to test an implication of the casual feature of the outcome approach. According to the outcome approach, frequency data are not considered to be as reliable a source in predicting outcomes as are phenomena that are casually related to the outcome. This being the case, it was predicted that outcome-oriented individuals would hold that if the casual features of a setup were altered, outcome frequencies for that setup would change accordingly. In the Modeling Problem, subjects were asked if it would be possible to construct an urn model of the bone that could be used to generate results that could not be distinguished from results obtained from rolling the bone. Subjects were introduced to the modeling concept in the Painted-die Problem where it was suggested that randomly sampling with replacement from an urn containing six identically-shaped stones would be the same as rolling a fair die. It was assumed that most subjects would accept this comparison since the most obvious physical feature -- the symmetry of the six sides -- was maintained. With an urn model of the bone, however, the important physical aspects of the bone -- its irregularly-shaped sides and unequal distribution of weight -- are transformed into unequal numbers of objects that are identical in weight and shape. It was predicted that outcome-oriented subjects, focusing on this difference, would expect that the data obtained from conducting trials on the two setups would be distinguishable in some way.

Scores for the Modeling Problem could range from 0 to 4, according to individual performance with respect to the following four categories:

(a) the urn model was not accepted in the case of the die;

(b) an urn model for rolling the bone could not be generated;

(c) a can filled with labeled stones corresponding in number to the statistician's estimates for each side was not accepted as a model of the bone;

(d) it was believed that no model of the bone could be created.

Interrater reliability for coding in these four categories was $r = .93$. The correlation between these scores and the outcome scores from Interview 1 was $r = .508$, $p < .05$ (one-tailed).

The reasons given by subjects for rejecting the urn models are congruent with the hypothesis that, in their analysis, important casual features could not be duplicated in the urn models. Subjects 3 and 13 stated that the urn model was inappropriate in the Painted-die Problem. They did not express concern over the corresponding features of the die and stone-filled urn per se, but over the differing sampling procedures in the two cases:

> S3: I think maybe the white side of the die would come up more, just 'cause you don't have any control over that [makes an imaginary roll of the die] -- Well, not that you do with the pieces....You're putting your hand in there and taking out. I just, I don't know why, but I don't think you'd pick the white one as often as the white side of the die.

> S13: I just think grabbing something out -- if you're grabbing it out, I think it would be more probably of being white. I don't know exactly why I'm thinking that way, but with this [die] I just [rolls die] -- I don't know, tossing something just seems less of a chance, but picking something out seems more of a chance. You'd think it would be the other way around, though. But I don't know...

In the following excerpts, subjects explain why an urn model is inappropriate in the case of the bone. The fact that the bone has six sides, uneven surfaces, and is rolled rather than drawn from are all mentioned as important differences between it and an urn filled with labeled objects.

S3: Probably be more likely to get no E's with the container full of 100 pieces. Just -- well, there is a slighter chance that it would come up, and there's six sides. So that's why I think it's more likely to come up on the bone.

I: Because?

S3: Because there's only six sides...

S6: Probably it would be more likely to get no E's from the bone, 'cause the bone has to stand like that, and it would be easier just sitting in there -- they don't have to -- it's not like there's anything to do with the way it can stand and stuff like that.

S6: [D] might be more likely from the bone. I don't really think you can say, but it just might be just because the D's are all mixed up in the can, whereas in the bone, that's the easiest side for it to land on. That's the most -- that's the way it stands easiest, so you might get it more times in a row in the bone.

S7: You could easily pick up 100 of them out without hitting an E. You'd have more trouble tossing the bone so you didn't come up with an E.

I: And why is that again?

S7: It just seems like because you're picking them out you could just miss one of the E's.

S15: These stones and the die are uniform, and each side is the same - it's the same surface. And this [bone] is all different. So this will affect -- the shape of the side will affect the way it's going to roll. Like it would be harder for it to stand up on E like that. So you'd have to replicate the little indents and stuff like -- So you couldn't make a -- you couldn't turn it into six stones or something like that.

The persistence demonstrated by students in insisting that the bone could not be modeled was particularly impressive. The interview probes were designed to give subjects several opportunities to accept a model: They were given one alternative after another. The independent coder, not knowing the intention in this probing, discreetly noted in two instances that the subjects had been strongly led to accept a model. The other subjects were as strongly "led" but insisted repeatedly that the model suggested would not be comparable to rolling the bone. Attending to the physical features as opposed to the resultant frequency data of a chance setup appears to be a deeply ingrained orientation.

General Discussion

As mentioned in the introduction, it has been suggested that two types of cognitions are available to adults in reasoning about uncertainty. These are (a) formal knowledge of probability theory and (b) natural assessments that become organized as judgment heuristics. Nisbett et al. (1983) have suggested that most adults will use formal, probabilistic knowledge when reasoning about situations that are clearly probabilistic and have a simple sample space. For situations that are less obviously probabilistic or for which the sample space is less tractable, they will fall back on the use of judgment heuristics.

Hidden in the above account is the assumption that regardless of whether the individual uses heuristics or formal probability knowledge, the individual perceives as the goal arriving at the probability of the event in question. While the value that is finally arrived at may be non-normative, the meaning of the value is assumed to lie somewhere in the range of acceptable interpretation.

The results of this study suggest that the above account is not complete. Many subjects who appeared to understand basic probability facts nevertheless could not apply this information to the fairly straightforward setup of the Painted-die Problem. Nor, alternatively, did they employ commonly-used judgment heuristics. It has been argued in this paper that these subjects approach uncertainty deterministically. This non-standard interpretation, labeled the outcome approach, is based on the objective of predicting outcomes of single trials.

When requested, outcome-oriented individuals will attach numeric values to their predictions. In this respect the outcome approach is similar to the personalist interpretation, in that the value associated with the prediction appears to be a measure of degree of belief. However, the similarity ends there. Personalist interpretations have been motivated by the desire to put subjective probabilities on a rational and scientific basis. Thus, among other requirements in these systems, subjective probabilities of repeated events should, over a long series of observed trials, closely approximate the actual frequencies of occurrence:

> If a person assesses the probability of a proposition being true as .7 and later finds that the proposition is false, that in itself does not invalidate the assessment. However, if a judge assigns .7 to 10,000 independent propositions, only 25 of which subsequently are found to be true, there is something wrong with these assessments. The attribute that they lack is called calibration.... Formally, a judge is calibrated if, over the long run, for all propositions assigned a given probability, the proportion that is true equals the probability assigned (Lichtenstein, Fischhoff, & Phillips, 1981, p. 306-307).

The outcome-oriented individual appears uninterested in calibration as defined above, but rather is interested in whether or not, on a particular occasion, a "correct" prediction can be made. If a non-predicted result occurs, the prediction was wrong and the confidence value, if assigned, was too high.

Another, and related, difference between the outcome approach and personalist interpretation is in the treatment of frequency information. Since a goal in a personalist interpretation is to be calibrated, the frequency of past occurrences of some event, when available, is used to formulate or adjust the initial probability. In the outcome approach, frequency data are not directly used to formulate confidence. It is especially clear in the Cab, Painted-die and Bone Problems that frequency information, when considered, is first translated into a more qualitative belief from which a numeric confidence can be subsequently generated if it is requested. A similar two-stage process of generating subjective probabilities has been suggested by Adams and Adams (1961) and more recently by Koriat, Lichtenstein and Fischhoff (1980).

> To assess one's confidence in the truth of a statement, one first arrives at a confidence judgment based on internal cues or "feelings of doubt".... The judgment is then transformed into a quantitative expression, such as a probability that the statement is correct (Koriat et al., p. 108).

It should be added that the latter step of quantifying internal cues is probably not an essential component of the outcome approach outside the laboratory. It seems to be done, and often begrudgingly, only if a request for a percentage or probability is made. In the outcome approach, discriminating between small differences in the strength of these inner feelings is unnecessary. Given the goal of predicting the most likely outcome on a particular occasion, one only need be aware of which outcome is associated with the strongest inner feeling. It is difficult to imagine, in fact, how quantifying one's confidence could aid the decision-making demands of most day-to-day situations.

On the other hand, not being able to translate from relevant quantitative information into belief strength is surely a handicap. Two possible reasons for this reluctant use of frequency data were previously mentioned -- that they are viewed as an unstable source of evidence and cannot be causally related to future events. Given only frequencies of past occurrence to predict future occurrence, it would seem that the prediction would of necessity reflect the uncertainty represented in the distribution of past occurrences. But the outcome-oriented individual apparently has not accepted uncertainty as inherent in certain domains. Subjects may even believe that someone who has mastered the mathematics of probability can predict the successive results of rolling a bone. As Subject 9 responded, "If I were a math major this would be easy."

Rather than frequency information, outcome-oriented subjects base predictions on data that are deterministically linked to the event of interest. The importance of causality in making judgments under uncertainty has been demonstrated in a variety of contexts. Azjen (1977), Nisbett and Ross (1980), Tversky and Kahneman (1980) and others have demonstrated that distributional information is more likely to be incorporated into probability estimates if presented in a way that strongly implies a causal link between features related to the data and the event of interest. Similar to performance on the Misfortune Problem, subjects given biographies of deviants tend to reconstruct the information so that the plight of the "victim" can be viewed as an inevitable result of life-events (Rosenhan, 1973). Also, subjects given descriptions of accidents search for a pattern in the associated events that make the accident appear predictable and avoidable (Walster, 1967). The betting behavior of professional gamblers as well as the way in which they toss dice suggests that they believe that they are controlling outcomes of chance events (Goffman, 1967).

If the outcome approach is a valid description of the novice's orientation to uncertainty, then the application of a causal rather than a black box model to uncertainty seems the most profound difference between the novice and the expert of probability, and thus, the most important to address in instruction. As long as students believe that there is some way that they can "know for sure" whether a specific hypothesis is correct, the better part of statistical logic and all of probability theory will evade them.

However, the preference for causal over stocastic models has, in this study, been linked to the preference for predicting outcomes of single trials rather than sample results. As Kahneman and Tversky have conjectured, "people generally prefer the singular mode, in which they take an 'inside view' of the causal system that most immediately produces the outcome, over an 'outside view' which relates the case at hand to a sampling schema" (1982, p. 153). The fact that these two tendencies are not independent, but logically support one another may explain in part why probability, as taught in the classroom, seems so foreign and difficult to master for many. While the application of causal reasoning to stocastic processes may be the most blatant demonstration of their lack of understanding, it may be more fruitful to attempt first to get students to focus on predicting sample results as opposed to single outcomes, thereby motivating a distributional schema.

References

Adams, J. K., & Adams, P. A. (1961). Realism of confidence judgments. Psychological Review, 68, 33-45.

Azjen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. Journal of Personality and Social Pyschology, 35, 303-314.

de Finetti, B. (1972). Probability, induction and statistics. New York, NY: Wiley.

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal report as data. Cambridge, MA: MIT Press.

Goffman, E. (1967). Interaction ritual. New York, NY: Anchor.

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

Kahneman, D., & Tversky, A. (1972). On prediction and judgment. Oregon Institute Bulletin, 12(4).

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237 - 251.

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

Konold, C. E. & Well, A. D. (1981). Analysis and reporting of interview data. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. Journal of Experimental Psychology: Human Learning and Memory, 6, 107-118.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1981). Calibration of probabilities: The state of the art to 1980. In Kahneman,D., Slovic, P., & Tversky, A. (Eds.), Judgment under uncertainty: Heuristics and biases. New York, NY: Cambridge University Press.

von Mises, R. (1957). Probability, statistics and truth. London: Allen and Unwin.

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.

Nisbett, R., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. New Jersey: Prentice-Hall.

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

Piaget, K., & Inhelder, B. (1975). The origin of the idea of chance. New York, NY: Norton. (Originally published, 1951.)

Pollatsek, A., Konold, C., Well, A. D., & Lima, S. D. (1984). Beliefs underlying random sampling. Memory & Cognition, 12, 395-401.

Reichenbach, H. (1949). The theory of probability. Los Angeles: University of California Press.

Rosenhan, D. (1973). On being sane in insane places. Science, 79, 250-252.

Savage, L. J. (1954). The foundation of statistics. New York, NY: Wiley.

Todhunter, I. (1949). A history of the mathematical theory of probability from the time of Pascal to that of LaPlace. New York: Chelsea. (Originally published, 1865).

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. Psychological Review, 76, 105-110.

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

Tversky, A., & Kahneman, D. (1980). Causal schemata in judgment under uncertainty. In Fishbein, M. (Ed.), Progress in social psychology. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunctive fallacy in probability judgment. Psychological Review, 90, 293-315.

Walster, E. (1967). "Second-guessing" important events. Human Relations, 20, 239-250.

Well, A. D., Pollatsek, A., & Konold, C. E. (1983). Probability estimation and the use and neglect of base-rate information. Unpublished manuscript, University of Massachusetts, Amherst.

Appendix

Problem Interview 1

Weather Problem

What does it mean when a weather forecaster says that tomorrow there is a 70% chance of rain? What does the number, in this case the 70%, tell you? How do they arrive at a specific number?

Suppose the forecaster said that there was a 70% chance of rain tomorrow and, in fact, it didn't rain. What would you conclude about the statement that there was a 70% chance of rain?

Suppose you wanted to find out how good a particular forecaster's predictions were. You observed what happened on ten days for which a 70% chance of rain had been reported. On three of those ten days there was no rain. What would you conclude about the accuracy of this forecaster? If the forecaster had been perfectly accurate, what would have happened? What should have been predicted on the days it didn't rain? With what percent chance?

Misfortune Problem

I know a person to whom all of the following things happened on the same day. First, his son totalled the family car and was seriously injury. Next, he was late for work and nearly got fired. In the afternoon he got food poisoning at a fast-food restaurant. Then in the evening he got word that his father had died. How would you account for all these things happening on the same day?

Bone Problem

I have here a bone that has six surfaces. I've written the letters A through F, one on each surface. (Subject is handed the bone which is labeled A, B, C, and D on the surfaces around the long axis, and E and F on the two surfaces at the ends of the long axis.) If you were to roll that, which side do you think would most likely land upright? How likely is it that $\underline{x}$ will land upright? (Subject is asked to roll the bone to see what happens.) What do you conclude about your prediction? What do you conclude having rolled the bone once? Would rolling the bone more times help you conclude which side is most likely to land upright?

(Subject is asked to roll the bone as many times as desired.) What do you conclude having rolled the bone several times? How many times would you have to roll the bone before you were absolutely confident about which side is most likely to land upright?

One day I got ambitious and rolled the bone 1000 times and recorded the results. This is what I got. (Subject is handed the list which showed A-50, B-279, C-244, D-375, E-52, F-0.) What do you conclude looking at these? Would you be willing to conclude that D is more likely than B? That B is more likely than C? That E is more likely than A? If asked what the chance was of rolling a D, what would you say?

I'm going to ask you to roll the bone ten times, but before you do, to predict how many of each side you will get. How did you arrive at those specific values? (Subject rolls the bone and records the results of each trial. After the 8th trial, the subject is asked:) What is your best guess of what you will get on the next two rolls? (After the last trial, the subject is asked:) How do you feel about your predictions? If you were going to roll the bone ten more times, what would you predict that you would get?

Problems: Interview 2

Cab Problem

(Subject is asked to read the Cab Problem aloud.) "A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

(i) 85% of the cabs in the city are Green, and 15% are Blue.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs, half of which were Blue and half of which were Green, the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

What is the probability that the cab involved in the accident was Blue rather than Green? (After subject gives a numerical response:) How did you arrive at that number?

Suppose the information in (i) were reversed such that 85% of the cabs in the city were Blue and 15% were Green. The witness, as before, identified it as Blue and was 80% correct in the test situation. In that case, what would you say the probability was that the cab involved in the accident was Blue?

Bone-2 Problem

Last time you were asked which side of this bone you thought would most likely land upright. Do you remember which side you concluded? (The bone is held far enough away so that the labels cannot be read.) I'm going to ask you the same question again. And to give you something to base your answer on, I'll offer you any one of the following pieces of information. (Subject is shown the list as the interviewer reads the items.)

1-A measure of surface area of each side.
2-The results of 100 rolls made by 16 people.
3-The results I got in 100C rolls.
4-A drawing of the bone showing the center of gravity.
5-The bone to look at.
6-The results on your last 10 rolls.

Which one would you like? Why did you choose that? If you could have a second piece of information, which would you choose? Why did you choose that?

(Subjects are given both choices unless item 4 has been picked. In that case, they are told that the drawing is not available, and to pick another item. The estimate of surface area is in square inches: A-.028, B-.078, C-.065, D-.169, E-.018, F-.031. The results of 100 rolls were: A-7, B-32, C-21, D-35, E-5, and F-0.) If you rolled the bone, which side do you think would most likely land upright? (Subject is asked to predict the results of ten trials, then the trials are conducted as in Interview 1.)

Painted-die Problem

I have here a six-sided die. Suppose I told you that there was a possibility that it was loaded -- that it had been altered so that one side was slightly more likely than the others to come up. Could you determine whether or not it was loaded? How? Would rolling it help you determine whether it was loaded? Suppose you rolled it 24 times and got the following results: (Subject is shown the results as the interviewer reads them.) 1-5, 2-2, 3-8, 4-2, 5-4, 6-3. What would you conclude?

In fact, the die is not loaded. Suppose I painted five of the surfaces black and the other one white. If I rolled the painted die six times, would I be more likely to get six blacks or five blacks and one white? If I rolled it 60 times, how many times would you expect the white surface to come up? (This probe was originally worded, "On the average, how many times would you have to roll the die until you got a white? After the third interview, it was changed to the present form, which was easier for subjects to understand.)

Obviously I haven't painted the die. But I do have five black stones and one white one. (The stones were identically shaped pieces from a board game.) Suppose I put these in this cup and shook it really well. Then I reached in without looking and drew one out, wrote down the color, replaced it, shook it up again and kept drawing like that. (This is demonstrated as it is explained.) Would that be the same as rolling the painted die? If I rolled the die several times and recorded what I got, and I drew stones and recorded those results, could you tell from looking at the results which I got from rolling the die and which from drawing stones? I'm going to draw six stones from the cup, but first ask you to predict what I'll get? (Stones are sampled, and before shown the results of each trial, the subject is asked both to predict the color that has been drawn, and the probability that it is that color.)

Modeling Problem

You agreed that we could create a model of the painted die by drawing stones from a certain cup -- that that would give comparable results. Would there be a similar way that we could make a model of the bone so that instead of rolling the bone, we could pick something out of a container and get the same kind of results?

(Subject is given the following probes successively until a model is agreed upon or the end of the list is reached:)

1- How about if we put six stones which have been labeled A through F in this cup and sampled from it as we did before?

2- Is there some container that I could fill with some number of lettered stones that would give results similar to rolling the bone?

3- Suppose we took the bone to a statistician and, however it is done, the following probabilities were calculated for each side: (Subject is shown the list as the interviewer reads it.) A was 5 out of 100, or 5%; B was 29 out of 100 or 29%; C, 24; D, 37; E, 5; and F, 0. So, we took a big can and first put five of these stones which have been labeled A inside. (A large can and six small containers filled with labeled stones are placed in front of the subject.) Then we took 29 B's, 24 C's, 37 D's, and 5 E's, and put them in the container. Then we shook it up and sampled from it as before. Do you think that would give results comparable to rolling the bone?

4- Suppose we rolled the bone and, say, we got B. We took a stone labeled B and put it in the container. Then we rolled the bone again, and similarly, whatever we got, we put the appropriately labeled stone in the container, and we did that over and over. Would we reach a point when it would make no difference if we rolled the bone or drew from the container we had filled?

(When, and if, the subject agreed upon a model of the bone, the following questions were asked:) Suppose I rolled the bone 100 times and kept track of what I got. Then I drew 100 times from this can filled with the labeled stones. If I showed you the results from both, could you tell from looking at the results which I got from rolling the bone and which from drawing from the container? In the 100 trials with the bone and the container, do you think with one of those I'd be more likely than with the other to get no E's? Do you think I'd be more likely with one of those to get more D's in 100 trials than with the other?

Table 1

Comparison of Outcome to Frequentist Responses

	Outcome Approach	Frequency Interpretation
	Weather Problem	
	I: What does it mean when a weather forecaster says that tomorrow there is a 70% chance of rain?	
(1)	S5: What it means is they can see all these cloud patterns forming and moving into a particular area, but they're not as dense as, say, a hurricane where you can absolutely predict where it's going to go. 100%--that means it was a total cloud thing coming over the area.	S4: 70% means that the chances that it will rain are seven out of ten, according to him.
	I: What does the number, in this case the 70%, tell you?	
(2)	S6: Well, it tells me that it's over 50%, and so, that's the first thing I think of. And, well, I think of the half-way mark between 50% and, say, 100% to be like, well, 75%. And it's almost that, and I think that's a pretty good chance that there'll be rain.	S4: W ll, it says that there's a 30% chance that it isn't going to rain.
	I: Suppose the forecaster said there was a 70% chance of rain tomorrow and, in fact, it didn't rain the next day. What would you conclude about the statement that there was a 70% chance of rain?	
(3)	S12: Well, that maybe they just fouled up. Or during the night, the precipitation or something changed in a different direction because of other outside factors.	S4: Well, on the basis of just the sample, I think an unrational response would be that the prediction was wrong. But, in fact, 30% is a pretty good probability that it's--it's not miniscule that it's not going to rain.

Table 1 (cont.)

I: Suppose you wanted to find out how good a particular forecaster's predictions were. You observed what happened on ten days for which a 70% chance of rain had been reported. On three of those ten days there was no rain. What would you conclude about the accuracy of this forecaster?

(4) S3: Well, I suppose he probably should do better than that. I assume they're trying their best. They're not trying to feed you wrong information.

S2: He was exactly right. Seven out of ten times is 70%. And he concluded 70% chance of rain all ten times. So--70% of all the time.

I: What should have been predicted on the days it didn't rain?

(5) S12: Well, he could either have said that there's a chance that it might rain rather than being more definite, or just said "mild," you know, "some clouds," or something like that rather than being specific.

Misfortune

I: I know of a person to whom all of the following things happened on the same day.... How would you account for all these things happening on the same day?

(6) S5: I'm trying to figure out if the order you gave me was the order that they happened, or if his father died--or he went out to a family restaurant with his family and they got food poisoning, and because he was sick, while he was driving he smashed up the car. His father died in the accident, and he was on his way to work so he was late.

S2: It's arbitrary, somewhat. It just occurred. I don't see any other way I could explain how they all occurred on the same day. I could see how if the guy totalled his car, he'd probably be late for work. Even though it's unlikely to occur, like if it only happens 1 in 1000 times, if you live 1000 days the odds are it's going to happen to you. So even though it's unlikely for an everyday occurrence, when you consider all the days that you live, it's not so unlikely.

Table 1 (cont.)

Bone Problem

I: If you were to roll this, which side do you think would most likely land upright?

(7) S9: Wow. If I were a math major, this would be easy. B is nice and flat, so if D fell down, B would be up. I'd go with B.

S2: I don't think I could tell you without rolling it. This is not like a die, and I think that there is no way of knowing personally without experimentation.

S4: I could only give my best guess. I'd have to say B up.

I: And about how likely do you think B is to land upright?

(8) S9: I wouldn't say it's much more likely. It depends on the roll, I think.

S4: I'll give a big bias to B. I'll say 33%.

I: So what do you conclude, having rolled it once?

(9) S10: Wrong again. [B] didn't come up.

S15: I don't conclude anything. Can I roll it again?

I: Would rolling it more times help you conclude which side, if any, was most likely to land upright?

(10) S9: No, I don't know. I think it's difficult to decide which is more likely. I don't see how you really can, just by looking at it. That's my opinion.

S1: Oh definitely. I mean that's the only way I could tell for sure. I think the only way with a thing like this is to just keep rolling it and jus record the results.

Table 2

Outcome-Oriented Responses: Interview 1

Problem/ Statement Description	1*	2*	3	4*	5	6	7	8*	9	10	11	12	13	14*	15	16*
	Subject Number															
Single trial feature: Evaluative response																
Bone: Prediction, right/wrong	X		X		X				X	X	X	X	X			X
Weather: Forecaster right/wrong			X						X		X	X	X		X	
7/10 -> pretty accurate			X	C[a]	C	X		C	X	X	C	X	X	X	X	X
Single trial feature: Qualitative interpretation:																
Weather: 50% < 70% < 100%	X		X		X	X		X	X			X	X	X	X	X
70% -> rain					X							X	X		X	
50% -> anything can ·en								X	X					X		
Casual feature																
Bone: Additional trials no help						X			X	X	X	X				
Ignore data of n=1000			X									X	X			
Predict via physical features						X			X			X				
Variability due to "the roll"			X		X			X	X		X	X	X		X	
Weather: 70% -> strength of causes			X						X			X				
No rain-> change of weather								X	X			X				
Misfortune: No mention of chance			X								X		X			
External, controlling force			X				X				X	X	X			
Internal, casual connection					X		X		X		X	X		X		
Outcome Score	2	0	9	0	6	4	2	4	11	3	7	13	9	4	5	3

* College or high school statistics course.

a C indicates conflict between "good" and "perfect" accuracy.

Table 3

Outcome-Oriented Responses: Interview 2

Problem/Category	Outcome Score											
	0	2	2	3	4	4	5	6	7	9	9	13
Cab Problem												
a) Number inquiry		X					X	X	X		X	
b) Qualitative statement		X					X					X
c) "Base" numeric answer										X	X	X
Bone-2 Problem												
a) 1st choice not frequency												X
b) 2nd choice not frequency			X				X	X		X		X
c) Predicted from physics properties								X		X		X
d) Statisticians use physical properties							X	X	X	X	X	X
Painted-die Problem												
a) Six blacks in six trials			X	X			X		X	X	X	X
b) <10 whites per 60 trials			X	X			X			X	X	X
c) p(B) > 5/6							X			X	X	X
Modeling Problem												
a) Reject urn model of die										X	X	
b) Urn model of bone not generated			X		X		X			X	X	X
c) Reject urn model of bone			X		X		X			X	X	X
d) No urn model of bone possible					X					X	X	
Subject Number	2	1	7	16	6	8	15	5	11	3	13	12