

3

Probabilistic Reasoning

Amos Tversky and Daniel Kahneman

Judgment under Uncertainty: Heuristics and Biases

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . .," "chances are . . .," "it is unlikely that . . .," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the probability of an uncertain event or the value of an uncertain quantity? This article shows that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

The subjective assessment of probability resembles the subjective assessment of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in the estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases.

First half from A. Tversky and D. Kahneman, Judgment under uncertainty: Heuristics and biases, *Science* 185, 1124–1131 (1974). Copyright 1974 by the AAAS. Second half from A. Tversky and D. Kahneman, Extensional versus intuitive reasoning: The conjunction fallacy in probabilistic reasoning, *Psychological Review* 90, 293–315 (1983). Copyright 1983 by the American Psychological Association. Reprinted by permission.

Such biases are also found in the intuitive judgment of probability. This article describes three heuristics [one omitted here—Ed.] that are employed to assess probabilities and to predict values. Biases to which these heuristics lead are enumerated, and the applied and theoretical implications of these observations are discussed.

Representativeness

Many of the probabilistic questions with which people are concerned belong to one of the following types: What is the probability that object A belongs to class B? What is the probability that event A originates from process B? What is the probability that process B will generate event A? In answering such questions, people typically rely on the representativeness heuristic, in which probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. For example, when A is highly representative of B, the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low.

For an illustration of judgment by representativeness, consider an individual who has been described by a former neighbor as follows: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." How do people assess the probability that Steve is engaged in a particular occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (Kahneman and Tversky, 1973). This approach to the judgment of probability leads to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.

Insensitivity to Prior Probability of Outcomes One of the factors that have no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes. In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into any reasonable estimate of the probability that Steve is a librarian rather than a farmer. Considerations of base-rate frequency, however, do not affect the similarity of Steve to the stereotypes of librarians and farmers. If people evaluate probability by representativeness, therefore, prior probabilities will be neglected. This hypothesis was tested in an exper-

iment where prior probabilities were manipulated (Kahneman and Tversky, 1973). Subjects were shown brief personality descriptions of several individuals, allegedly sampled at random from a group of 100 professionals—engineers and lawyers. The subjects were asked to assess, for each description, the probability that it belonged to an engineer rather than to a lawyer. In one experimental condition, subjects were told that the group from which the descriptions had been drawn consisted of 70 engineers and 30 lawyers. In another condition, subjects were told that the group consisted of 30 engineers and 70 lawyers. The odds that any particular description belongs to an engineer rather than to a lawyer should be higher in the first condition, where there is a majority of engineers, than in the second condition, where there is a majority of lawyers. Specifically, it can be shown by applying Bayes' rule that the ratio of these odds should be $(.7/.3)^2$, or 5.44, for each description. In a sharp violation of Bayes' rule, the subjects in the two conditions produced essentially the same probability judgments. Apparently, subjects evaluated the likelihood that a particular description belonged to an engineer rather than to a lawyer by the degree to which this description was representative of the two stereotypes, with little or no regard for the prior probabilities of the categories.

The subjects used prior probabilities correctly when they had no other information. In the absence of a personality sketch, they judged the probability that an unknown individual is an engineer to be .7 and .3, respectively, in the two base-rate conditions. However, prior probabilities were effectively ignored when a description was introduced, even when this description was totally uninformative. The responses to the following description illustrate this phenomenon:

Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

This description was intended to convey no information relevant to the question of whether Dick is an engineer or a lawyer. Consequently, the probability that Dick is an engineer should equal the proportion of engineers in the group, as if no description had been given. The subjects, however, judged the probability of Dick being an engineer to be .5 regardless of whether the stated proportion of engineers in the group was .7 or .3. Evidently, people respond differently when given no evidence and when given worthless evidence. When no specific evidence is given, prior probabilities are properly utilized; when worthless evidence is given, prior probabilities are ignored (Kahneman and Tversky 1973).

Insensitivity to Sample Size To evaluate the probability of obtaining a particular result in a sample drawn from a specified population, people typically apply the representativeness heuristic. That is, they assess the

likelihood of a sample result, for example, that the average height in a random sample of ten men will be 6 feet (180 centimeters), by the similarity of this result to the corresponding parameter (that is, to the average height in the population of men). The similarity of a sample statistic to a population parameter does not depend on the size of the sample. Consequently, if probabilities are assessed by representativeness, then the judged probability of a sample statistic will be essentially independent of sample size. Indeed, when subjects assessed the distributions of average height for samples of various sizes, they produced identical distributions. For example, the probability of obtaining an average height greater than 6 feet was assigned the same value for samples of 1000, 100, and 10 men (Kahneman and Tversky 1972). Moreover, subjects failed to appreciate the role of sample size even when it was emphasized in the formulation of the problem. Consider the following question:

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower.

For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?

The larger hospital (21)

The smaller hospital (21)

About the same (that is, within 5 percent of each other) (53)

The values in parentheses are the number of undergraduate students who chose each answer.

Most subjects judged the probability of obtaining more than 60 percent boys to be the same in the small and in the large hospital, presumably because these events are described by the same statistic and are therefore equally representative of the general population. In contrast, sampling theory entails that the expected number of days on which more than 60 percent of the babies are boys is much greater in the small hospital than in the large one, because a large sample is less likely to stray from 50 percent. This fundamental notion of statistics is evidently not part of people's repertoire of intuitions.

A similar insensitivity to sample size has been reported in judgments of posterior probability, that is, of the probability that a sample has been drawn from one population rather than from another. Consider the following example:

Imagine an urn filled with balls, of which $\frac{2}{3}$ are of one color and $\frac{1}{3}$ of another. One individual has drawn 5 balls from the urn, and found that 4 were red and 1 was white. Another individual has drawn 20 balls and found that 12 were red and 8 were white. Which of the two individuals should feel more confident that the urn contains $\frac{2}{3}$ red balls

and $\frac{1}{3}$ white balls, rather than the opposite? What odds should each individual give?

In this problem, the correct posterior odds are 8 to 1 for the 4:1 sample and 16 to 1 for the 12:8 sample, assuming equal prior probabilities. However, most people feel that the first sample provides much stronger evidence for the hypothesis that the urn is predominantly red, because the proportion of red balls is larger in the first than in the second sample. Here again, intuitive judgments are dominated by the sample proportion and are essentially unaffected by the size of the sample, which plays a crucial role in the determination of the actual posterior odds (Kahneman and Tversky, 1972). In addition, intuitive estimates of posterior odds are far less extreme than the correct values. The underestimation of the impact of evidence has been observed repeatedly in problems of this type (W. Edwards, 1968; Slovic and Lichtenstein 1971). It has been labeled "conservatism."

Misconceptions of Chance People expect that a sequence of events generated by a random process will represent the essential characteristics of that process even when the sequence is short. In considering tosses of a coin for heads or tails, for example, people regard the sequence H-T-H-T-T-H to be more likely than the sequence H-H-H-T-T-T, which does not appear random, and also more likely than the sequence H-H-H-H-T-H, which does not represent the fairness of the coin (Kahneman and Tversky, 1972). Thus, people expect that the essential characteristics of the process will be represented, not only globally in the entire sequence, but also locally in each of its parts. A locally representative sequence, however, deviates systematically from chance expectation: it contains too many alternations and too few runs. Another consequence of the belief in local representativeness is the well-known gambler's fallacy. After observing a long run of red on the roulette wheel, for example, most people erroneously believe that black is now due, presumably because the occurrence of black will result in a more representative sequence than the occurrence of an additional red. Chance is commonly viewed as a self-correcting process in which a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium. In fact, deviations are not "corrected" as a chance process unfolds, they are merely diluted.

Misconceptions of chance are not limited to naive subjects. A study of the statistical intuitions of experienced research psychologists (Tversky and Kahneman 1971) revealed a lingering belief in what may be called the "law of small numbers," according to which even small samples are highly representative of the populations from which they are drawn. The responses of these investigators reflected the expectation that a valid hypothesis about a population will be represented by a statistically significant result in a sample—with little regard for its size. As a consequence, the researchers put too much faith in the results

of small samples and grossly overestimated the replicability of such results. In the actual conduct of research, this bias leads to the selection of samples of inadequate size and to overinterpretation of findings.

Insensitivity to Predictability People are sometimes called upon to make such numerical predictions as the future value of a stock, the demand for a commodity, or the outcome of a football game. Such predictions are often made by representativeness. For example, suppose one is given a description of a company and is asked to predict its future profit. If the description of the company is very favorable, a very high profit will appear most representative of that description; if the description is mediocre, a mediocre performance will appear most representative. The degree to which the description is favorable is unaffected by the reliability of that description or by the degree to which it permits accurate prediction. Hence, if people predict solely in terms of the favorableness of the description, their predictions will be insensitive to the reliability of the evidence and to the expected accuracy of the prediction.

This mode of judgment violates the normative statistical theory in which the extremeness and the range of predictions are controlled by considerations of predictability. When predictability is nil, the same prediction should be made in all cases. For example, if the descriptions of companies provide no information relevant to profit, then the same value (such as average profit) should be predicted for all companies. If predictability is perfect, of course, the values predicted will match the actual values and the range of predictions will equal the range of outcomes. In general, the higher the predictability, the wider the range of predicted values.

Several studies of numerical prediction have demonstrated that intuitive predictions violate this rule, and that subjects show little or no regard for considerations of predictability (Kahneman and Tversky 1973). In one of these studies, subjects were presented with several paragraphs, each describing the performance of a student teacher during a particular practice lesson. Some subjects were asked to *evaluate* the quality of the lesson described in the paragraph in percentile scores, relative to a specified population. Other subjects were asked to *predict*, also in percentile scores, the standing of each student teacher 5 years after the practice lesson. The judgments made under the two conditions were identical. That is, the prediction of a remote criterion (success of a teacher after 5 years) was identical to the evaluation of the information on which the prediction was based (the quality of the practice lesson). The students who made these predictions were undoubtedly aware of the limited predictability of teaching competence on the basis of a single trial lesson 5 years earlier; nevertheless, their predictions were as extreme as their evaluations.

The Illusion of Validity As we have seen, people often predict by selecting the outcome (for example, an occupation) that is most representative of the input (for example, the description of a person). The confidence they have in their prediction depends primarily on the degree of representativeness (that is, on the quality of the match between the selected outcome and the input) with little or no regard for the factors that limit predictive accuracy. Thus, people express great confidence in the prediction that a person is a librarian when given a description of his personality which matches the stereotype of librarians, even if the description is scanty, unreliable, or outdated. The unwarranted confidence which is produced by a good fit between the predicted outcome and the input information may be called the illusion of validity. This illusion persists even when the judge is aware of the factors that limit the accuracy of his predictions. It is a common observation that psychologists who conduct selection interviews often experience considerable confidence in their predictions, even when they know of the vast literature that shows selection interviews to be highly fallible. The continued reliance on the clinical interview for selection, despite repeated demonstrations of its inadequacy, amply attests to the strength of this effect.

The internal consistency of a pattern of inputs is a major determinant of one's confidence in predictions based on these inputs. For example, people express more confidence in predicting the final grade-point average of a student whose first-year record consists entirely of B's than in predicting the grade-point average of a student whose first-year record includes many A's and C's. Highly consistent patterns are most often observed when the input variables are highly redundant or correlated. Hence, people tend to have great confidence in predictions based on redundant input variables. However, an elementary result in the statistics of correlation asserts that, given input variables of stated validity, a prediction based on several such inputs can achieve higher accuracy when they are independent of each other than when they are redundant or correlated. Thus, redundancy among inputs decreases accuracy even as it increases confidence, and people are often confident in predictions that are quite likely to be off the mark (Kahneman and Tversky 1973).

Misconceptions of Regression Suppose a large group of children has been examined on two equivalent versions of an aptitude test. If one selects ten children from among those who did best on one of the two versions, he will usually find their performance on the second version to be somewhat disappointing. Conversely, if one selects ten children from among those who did worst on one version, they will be found, on the average, to do somewhat better on the other version. More generally, consider two variables X and Y which have the same distribution. If one selects individuals whose average X score deviates from

the mean of X by k units, then the average of their Y scores will usually deviate from the means of Y by less than k units. These observations illustrate a general phenomenon known as regression toward the mean, which was first documented by Galton more than 100 years ago.

In the normal course of life, one encounters many instances of regression toward the mean, in the comparison of the height of fathers and sons, of the intelligence of husbands and wives, or of the performance of individuals on consecutive examinations. Nevertheless, people do not develop correct intuitions about this phenomenon. First, they do not expect regression in many contexts where it is bound to occur. Second, when they recognize the occurrence of regression, they often invent spurious causal explanations for it (Kahneman and Tversky 1973). We suggest that the phenomenon of regression remains elusive because it is incompatible with the belief that the predicted outcome should be maximally representative of the input, and hence, that the value of the outcome variable should be as extreme as the value of the input variable.

The failure to recognize the import of regression can have pernicious consequences, as illustrated by the following observation (Kahneman and Tversky 1973). In a discussion of flight training, experienced instructors noted that praise for an exceptionally smooth landing is typically followed by a poorer landing on the next try, while harsh criticism after a rough landing is usually followed by an improvement on the next try. The instructors concluded that verbal rewards are detrimental to learning, while verbal punishments are beneficial, contrary to accepted psychological doctrine. This conclusion is unwarranted because of the presence of regression toward the mean. As in other cases of repeated examination, an improvement will usually follow a poor performance and a deterioration will usually follow an outstanding performance, even if the instructor does not respond to the trainee's achievement on the first attempt. Because the instructors had praised their trainees after good landings and admonished them after poor ones, they reached the erroneous and potentially harmful conclusion that punishment is more effective than reward.

Thus, the failure to understand the effect of regression leads one to overestimate the effectiveness of punishment and to underestimate the effectiveness of reward. In social interaction, as well as in training, rewards are typically administered when performance is good, and punishments are typically administered when performance is poor. By regression alone, therefore, behavior is most likely to improve after punishment and most likely to deteriorate after reward. Consequently, the human condition is such that, by chance alone, one is most often rewarded for punishing others and most often punished for rewarding them. People are generally not aware of this contingency. In fact, the elusive role of regression in determining the apparent consequences of

reward and punishment seems to have escaped the notice of students of this area.

Availability

There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind. For example, one may assess the risk of heart attack among middle-aged people by recalling such occurrences among one's acquaintances. Similarly, one may evaluate the probability that a given business venture will fail by imagining various difficulties it could encounter. This judgmental heuristic is called availability. Availability is a useful clue for assessing frequency or probability, because instances of large classes are usually reached better and faster than instances of less frequent classes. However, availability is affected by factors other than frequency and probability. Consequently, the reliance on availability leads to predictable biases, some of which are illustrated below.

Biases Due to the Retrievability of Instances When the size of a class is judged by the availability of its instances, a class whose instances are easily retrieved will appear more numerous than a class of equal frequency whose instances are less retrievable. In an elementary demonstration of this effect, subjects heard a list of well-known personalities of both sexes and were subsequently asked to judge whether the list contained more names of men than of women. Different lists were presented to different groups of subjects. In some of the lists the men were relatively more famous than the women, and in others the women were relatively more famous than the men. In each of the lists, the subjects erroneously judged that the class (sex) that had the more famous personalities was the more numerous (Tversky and Kahneman 1973).

In addition to familiarity, there are other factors, such as salience, which affect the retrievability of instances. For example, the impact of seeing a house burning on the subjective probability of such accidents is probably greater than the impact of reading about a fire in the local paper. Furthermore, recent occurrences are likely to be relatively more available than earlier occurrences. It is a common experience that the subjective probability of traffic accidents rises temporarily when one sees a car overturned by the side of the road.

Biases Due to the Effectiveness of a Search Set Suppose one samples a word (of three letters or more) at random from an English text. Is it more likely that the word starts with *r* or that *r* is the third letter? People approach this problem by recalling words that begin with *r* (road) and words that have *r* in the third position (car) and assess the relative

frequency by the ease with which words of the two types come to mind. Because it is much easier to search for words by their first letter than by their third letter, most people judge words that begin with a given consonant to be more numerous than words in which the same consonant appears in the third position. They do so even for consonants, such as *r* or *k*, that are more frequent in the third position than in the first (Tversky and Kahneman 1973).

Different tasks elicit different search sets. For example, suppose you are asked to rate the frequency with which abstract words (*thought, love*) and concrete words (*door, water*) appear in written English. A natural way to answer this question is to search for contexts in which the word could appear. It seems easier to think of contexts in which an abstract concept is mentioned (*love* in love stories) than to think of contexts in which a concrete word (such as *door*) is mentioned. If the frequency of words is judged by the availability of the contexts in which they appear, abstract words will be judged as relatively more numerous than concrete words. This bias has been observed in a recent study (Galbraith and Underwood 1973) which showed that the judged frequency of occurrence of abstract words was much higher than that of concrete words, equated in objective frequency. Abstract words were also judged to appear in a much greater variety of contexts than concrete words.

Biases of Imaginability Sometimes one has to assess the frequency of a class whose instances are not stored in memory but can be generated according to a given rule. In such situations, one typically generates several instances and evaluates frequency or probability by the ease with which the relevant instances can be constructed. However, the ease of constructing instances does not always reflect their actual frequency, and this mode of evaluation is prone to biases. To illustrate, consider a group of 10 people who form committees of k members, $2 \leq k \leq 8$. How many different committees of k members can be formed? The correct answer to this problem is given by the binomial coefficient $\binom{10}{k}$ which reaches a maximum of 252 for $k = 5$. Clearly, the number of committees of k members equals the number of committees of $(10 - k)$ members, because any committee of k members defines a unique group of $(10 - k)$ nonmembers.

One way to answer this question without computation is to mentally construct committees of k members and to evaluate their number by the ease with which they come to mind. Committees of few members, say 2, are more available than committees of many members, say 8. The simplest scheme for the construction of committees is a partition of the group into disjoint sets. One readily sees that it is easy to construct five disjoint committees of 2 members, while it is impossible to generate even two disjoint committees of 8 members. Consequently, if frequency is assessed by imaginability, or by availability for construction, the small committees will appear more numerous than larger

committees, in contrast to the correct bell-shaped function. Indeed, when naive subjects were asked to estimate the number of distinct committees of various sizes, their estimates were a decreasing monotonic function of committee size (Tversky and Kahneman 1973). For example, the median estimate of the number of committees of 2 members was 70, while the estimate for committees of 8 members was 20 (the correct answer is 45 in both cases).

Imaginability plays an important role in the evaluation of probabilities in real-life situations. The risk involved in an adventurous expedition, for example, is evaluated by imaging contingencies with which the expedition is not equipped to cope. If many such difficulties are vividly portrayed, the expedition can be made to appear exceedingly dangerous, although the ease with which disasters are imagined need not reflect their actual likelihood. Conversely, the risk involved in an undertaking may be grossly underestimated if some possible dangers are either difficult to conceive of, or simply do not come to mind.

Illusory Correlation Chapman and Chapman (1969) have described an interesting bias in the judgment of the frequency with which two events co-occur. They presented naive judges with information concerning several hypothetical mental patients. The data for each patient consisted of a clinical diagnosis and a drawing of a person made by the patient. Later the judges estimated the frequency with which each diagnosis (such as paranoia or suspiciousness) had been accompanied by various features of the drawing (such as peculiar eyes). The subjects markedly overestimated the frequency of co-occurrence of natural associates, such as suspiciousness and peculiar eyes. This effect was labeled illusory correlation. In their erroneous judgments of the data to which they had been exposed, naive subjects "rediscovered" much of the common, but unfounded, clinical lore concerning the interpretation of the draw-a-person test. The illusory correlation effect was extremely resistant to contradictory data. It persisted even when the correlation between symptom and diagnosis was actually negative, and it prevented the judges from detecting relationships that were in fact present.

Availability provides a natural account for the illusory-correlation effect. The judgment of how frequently two events co-occur could be based on the strength of the associative bond between them. When the association is strong, one is likely to conclude that the events have been frequently paired. Consequently, strong associates will be judged to have occurred together frequently. According to this view, the illusory correlation between suspiciousness and peculiar drawing of the eyes, for example, is due to the fact that suspiciousness is more readily associated with the eyes than with any other part of the body.

Lifelong experience has taught us that, in general, instances of large classes are recalled better and faster than instances of less frequent classes; that likely occurrences are easier to imagine than unlikely ones;

and that the associative connections between events are strengthened when the events frequently co-occur. As a result, man has at his disposal a procedure (the availability heuristic) for estimating the numerosity of a class, the likelihood of an event, or the frequency of co-occurrences, by the ease with which the relevant mental operations of retrieval, construction, or association can be performed. However, as the preceding examples have demonstrated, this valuable estimation procedure results in systematic errors. . . .

The Conjunction Fallacy in Probabilistic Reasoning

. . . The laws of probability derive from extensional considerations. A probability measure is defined on a family of events and each event is construed as a set of possibilities, such as the three ways of getting a 10 on a throw of a pair of dice. The probability of an event equals the sum of the probabilities of its disjoint outcomes. Probability theory has traditionally been used to analyze repetitive chance processes, but the theory has also been applied to essentially unique events where probability is not reducible to the relative frequency of "favorable" outcomes. The probability that the man who sits next to you on the plane is unmarried equals the probability that he is a bachelor plus the probability that he is either divorced or widowed. Additivity applies even when probability does not have a frequentistic interpretation and when the elementary events are not equiprobable.

The simplest and most fundamental qualitative law of probability is the extension rule: If the extension of A includes the extension of B (i.e., $A \supset B$) then $P(A) \geq P(B)$. Because the set of possibilities associated with a conjunction $A \& B$ is included in the set of possibilities associated with B , the same principle can also be expressed by the conjunction rule $P(A \& B) \leq P(B)$: A conjunction cannot be more probable than one of its constituents. This rule holds regardless of whether A and B are independent and is valid for any probability assignment on the same sample space. Furthermore, it applies not only to the standard probability calculus but also to nonstandard models such as upper and lower probability (Dempster 1967; Suppes 1975), belief function (Shafer 1976), Baconian probability (Cohen 1977), rational belief (Kyburg 1983), and possibility theory (Zadeh 1978).

In contrast to formal theories of belief, intuitive judgments of probability are generally not extensional. People do not normally analyze daily events into exhaustive lists of possibilities or evaluate compound probabilities by aggregating elementary ones. Instead, they commonly use a limited number of heuristics, such as representativeness and availability (Kahneman et al. 1982). Our conception of judgmental heuristics is based on *natural assessments* that are routinely carried out as part of the perception of events and the comprehension of messages. Such natural assessments include computations of similarity and rep-

representativeness, attributions of causality, and evaluations of the availability of associations and exemplars. These assessments, we propose, are performed even in the absence of a specific task set, although their results are used to meet task demands as they arise. For example, the mere mention of “horror movies” activates instances of horror movies and evokes an assessment of their availability. Similarly, the statement that Woody Allen’s aunt had hoped that he would be a dentist elicits a comparison of the character to the stereotype and an assessment of representativeness. It is presumably the mismatch between Woody Allen’s personality and our stereotype of a dentist that makes the thought mildly amusing. Although these assessments are not tied to the estimation of frequency or probability, they are likely to play a dominant role when such judgments are required. The availability of horror movies may be used to answer the question, “What proportion of the movies produced last year were horror movies?,” and representativeness may control the judgment that a particular boy is more likely to be an actor than a dentist.

The term *judgmental heuristic* refers to a strategy—whether deliberate or not—that relies on a natural assessment to produce an estimation or a prediction. One of the manifestations of a heuristic is the relative neglect of other considerations. For example, the resemblance of a child to various professional stereotypes may be given too much weight in predicting future vocational choice, at the expense of other pertinent data such as the base rate frequencies of occupations. Hence, the use of judgmental heuristics gives rise to predictable biases. Natural assessments can affect judgments in other ways, for which the term *heuristic* is less apt. First, people sometimes misinterpret their task and fail to distinguish the required judgment from the natural assessment that the problem evokes. Second, the natural assessment may act as an anchor to which the required judgment is assimilated, even when the judge does not intend to use the one to estimate the other.

Previous discussions of errors of judgment have focused on deliberate strategies and on misinterpretation of tasks. The present treatment calls special attention to the processes of anchoring and assimilation, which are often neither deliberate nor conscious. An example from perception may be instructive: If two objects in a picture of a three-dimensional scene have the same picture size, the one that appears more distant is not only seen as “really” larger but also as larger in the picture. The natural computation of real size evidently influences the (less natural) judgment of picture size, although observers are unlikely to confuse the two values or to use the former to estimate the latter.

The natural assessments of representativeness and availability do not conform to the extensional logic of probability theory. In particular, a conjunction can be more representative than one of its constituents, and instances of a specific category can be easier to retrieve than instances of a more inclusive category. The following demonstration il-

illustrates the point. When they were given 60 sec to list seven-letter words of a specified form, students at the University of British Columbia (UBC) produced many more words of the form _ _ _ _ i n g than of the form _ _ _ _ n _ , although the latter class includes the former. The average numbers of words produced in the two conditions were 6.4 and 2.9, respectively, $t(44) = 4.70, p < .01$. In this test of availability, the increased efficacy of memory search suffices to offset the reduced extension of the target class.

Our treatment of the availability heuristic (Tversky and Kahneman 1973) suggests that the differential availability of *ing* words and of _ n _ words should be reflected in judgments of frequency. The following questions test this prediction.

In four pages of a novel (about 2,000 words), how many words would you expect to find that have the form _ _ _ _ i n g (seven-letter words that end with "ing")? Indicate your best estimate by circling one of the values below:

0 1-2 3-4 5-7 8-10 11-15 16+

A second version of the question requested estimates for words of the form _ _ _ _ n _ . The median estimates were 13.4 for *ing* words ($n = 52$), and 4.7 for _ n _ words ($n = 53, p < .01$, by median test), contrary to the extension rule. Similar results were obtained for the comparison of words of the form _ _ _ _ l y with words of the form _ _ _ _ l _ ; the median estimates were 8.8 and 4.4, respectively.

This example illustrates the structure of the studies reported in this article. We constructed problems in which a reduction of extension was associated with an increase in availability or representativeness, and we tested the conjunction rule in judgments of frequency or probability. In the next section we discuss the representativeness heuristic and contrast it with the conjunction rule in the context of person perception. The third section describes conjunction fallacies in medical prognoses, sports forecasting, and choice among bets. In the fourth section we investigate probability judgments for conjunctions of causes and effects and describe conjunction errors in scenarios of future events. Manipulations that enable respondents to resist the conjunction fallacy are explored in the fifth section, and the implications of the results are discussed in the last section. [The third through fifth sections are omitted.—Ed.]

Representative Conjunctions

Modern research on categorization of objects and events (Mervis and Rosch 1981; Rosch 1978; Smith and Medin 1981) has shown that information is commonly stored and processed in relation to mental models, such as prototypes and schemata. It is therefore natural and economical for the probability of an event to be evaluated by the degree to which that event is representative of an appropriate mental model (Kahneman and Tversky 1972, 1973; Tversky and Kahneman, 1971, 1982). Because

many of the results reported here are attributed to this heuristic, we first briefly analyze the concept of representativeness and illustrate its role in probability judgment.

Representativeness is an assessment of the degree of correspondence between a sample and a population, an instance and a category, an act and an actor or, more generally, between an outcome and a model. The model may refer to a person, a coin, or the world economy, and the respective outcomes could be marital status, a sequence of heads and tails, or the current price of gold. Representativeness can be investigated empirically by asking people, for example, which of two sequences of heads and tails is more representative of a fair coin or which of two professions is more representative of a given personality. This relation differs from other notions of proximity in that it is distinctly directional. It is natural to describe a sample as more or less representative of its parent population or a species (e.g., robin, penguin) as more or less representative of a superordinate category (e.g., bird). It is awkward to describe a population as representative of a sample or a category as representative of an instance.

When the model and the outcomes are described in the same terms, representativeness is reducible to similarity. Because a sample and a population, for example, can be described by the same attributes (e.g., central tendency and variability), the sample appears representative if its salient statistics match the corresponding parameters of the population. In the same manner, a person seems representative of a social group if his or her personality resembles the stereotypical member of that group. Representativeness, however, is not always reducible to similarity; it can also reflect causal and correlational beliefs (see, e.g., Chapman and Chapman 1967; Jennings et al. 1982; Nisbett and Ross 1980). A particular act (e.g., suicide) is representative of a person because we attribute to the actor a disposition to commit the act, not because the act resembles the person. Thus, an outcome is representative of a model if the salient features match or if the model has a propensity to produce the outcome.

Representativeness tends to covary with frequency: Common instances and frequent events are generally more representative than unusual instances and rare events. The representative summer day is warm and sunny, the representative American family has two children, and the representative height of an adult male is about 5 feet 10 inches. However, there are notable circumstances where representativeness is at variance with both actual and perceived frequency. First, a highly specific outcome can be representative but infrequent. Consider a numerical variable, such as weight, that has a unimodal frequency distribution in a given population. A narrow interval near the mode of the distribution is generally more representative of the population than a wider interval near the tail. For example, 68% of a group of Stanford University undergraduates ($N = 105$) stated that it is more representa-

tive for a female Stanford student “to weigh between 124 and 125 pounds” than “to weigh more than 135 pounds”. On the other hand, 78% of a different group ($N = 102$) stated that among female Stanford students there are more “women who weigh more than 135 pounds” than “women who weigh between 124 and 125 pounds.” Thus, the narrow modal interval (124–125 pounds) was judged to be more representative but less frequent than the broad tail interval (above 135 pounds).

Second, an attribute is representative of a class if it is very diagnostic, that is, if the relative frequency of this attribute is much higher in that class than in a relevant reference class. For example, 65% of the subjects ($N = 105$) stated that it is more representative for a Hollywood actress “to be divorced more than 4 times” than “to vote Democratic.” Multiple divorce is diagnostic of Hollywood actresses because it is part of the stereotype that the incidence of divorce is higher among Hollywood actresses than among other women. However, 83% of a different group ($N = 102$) stated that, among Hollywood actresses, there are more “women who vote Democratic” than “women who are divorced more than 4 times.” Thus, the more diagnostic attribute was judged to be more representative but less frequent than an attribute (voting Democratic) of lower diagnosticity. Third, an unrepresentative instance of a category can be fairly representative of a superordinate category. For example, chicken is a worse exemplar of a bird than of an animal, and rice is an unrepresentative vegetable, although it is a representative food.

The preceding observations indicate that representativeness is nonextensional: It is not determined by frequency, and it is not bound by class inclusion. Consequently, the test of the conjunction rule in probability judgments offers the sharpest contrast between the extensional logic of probability theory and the psychological principles of representativeness. Our first set of studies of the conjunction rule were conducted in 1974, using occupation and political affiliation as target attributes to be predicted singly or in conjunction from brief personality sketches (see Tversky and Kahneman 1982, for a brief summary). The studies described in the present section replicate and extend our earlier work. We used the following personality sketches of two fictitious individuals, Bill and Linda, followed by a set of occupations and avocations associated with each of them.

Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Bill is a physician who plays poker for a hobby.

Bill is an architect.

Bill is an accountant. (A)

Bill plays jazz for a hobby. (J)

Bill surfs for a hobby.

Bill is a reporter.

Bill is an accountant who plays jazz for a hobby. (A&J)

Bill climbs mountains for a hobby.

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Linda is a teacher in elementary school.

Linda works in a bookstore and takes Yoga classes.

Linda is active in the feminist movement. (F)

Linda is a psychiatric social worker.

Linda is a member of the League of Women Voters.

Linda is a bank teller. (T)

Linda is an insurance salesperson.

Linda is a bank teller and is active in the feminist movement. (T&F)

As the reader has probably guessed, the description of Bill was constructed to be representative of an accountant (A) and unrepresentative of a person who plays jazz for a hobby (J). The description of Linda was constructed to be representative of an active feminist (F) and unrepresentative of a bank teller (T). We also expected the ratings of representativeness to be higher for the classes defined by a conjunction of attributes (A&J for Bill, T&F for Linda) than for the less representative constituent of each conjunction (J and T, respectively).

A group of 88 undergraduates at UBC ranked the eight statements associated with each description by "the degree to which Bill (Linda) resembles the typical member of that class." The results confirmed our expectations. The percentages of respondents who displayed the predicted order ($A > A\&J > J$ for Bill; $F > T\&F > T$ for Linda) were 87% and 85%, respectively. This finding is neither surprising nor objectionable. If, like similarity and prototypicality, representativeness depends on both common and distinctive features (Tversky 1977), it should be enhanced by the addition of shared features. Adding eyebrows to a schematic face makes it more similar to another schematic face with eyebrows (Gati and Tversky 1982). Analogously, the addition of feminism to the profession of bank teller improves the match of Linda's current activities to her personality. More surprising and less acceptable is the finding that the great majority of subjects also rank the conjunctions (A&J and T&F) as more *probable* than their less representative constituents (J and T). The following sections describe and analyze this phenomenon.

Indirect and Subtle Tests Experimental tests of the conjunction rule can be divided into three types: *indirect* tests, *direct-subtle* tests and *direct-transparent* tests. In the indirect tests, one group of subjects evaluates the probability of the conjunction, and another group of subjects evaluates the probability of its constituents. No subject is required to com-

pare a conjunction (e.g., “Linda is a bank teller and a feminist”) to its constituents. In the direct-subtle tests, subjects compare the conjunction to its less representative constituent, but the inclusion relation between the events is not emphasized. In the direct-transparent tests, the subjects evaluate or compare the probabilities of the conjunction and its constituent in a format that highlights the relation between them.

The three experimental procedures investigate different hypotheses. The indirect procedure tests whether probability judgments conform to the conjunction rule; the direct-subtle procedure tests whether people will take advantage of an opportunity to compare the critical events; the direct-transparent procedure tests whether people will obey the conjunction rule when they are compelled to compare the critical events. This sequence of tests also describes the course of our investigation, which began with the observation of violations of the conjunction rule in indirect tests and proceeded—to our increasing surprise—to the finding of stubborn failures of that rule in several direct-transparent tests.

Three groups of respondents took part in the main study. The statistically *naive* group consisted of undergraduate students at Stanford University and UBC who had no background in probability or statistics. The *informed* group consisted of first-year graduate students in psychology and in education and of medical students at Stanford who were all familiar with the basic concepts of probability after one or more courses in statistics. The *sophisticated* group consisted of doctoral students in the decision science program of the Stanford Business School who had taken several advanced courses in probability, statistics, and decision theory.

Subjects in the main study received one problem (either Bill or Linda) first in the format of a direct test. They were asked to rank all eight statements associated with that problem (including the conjunction, its separate constituents, and five filler items) according to their probability, using 1 for the most probable and 8 for the least probable. The subjects then received the remaining problem in the format of an indirect test in which the list of alternatives included either the conjunction or its separate constituents. The same five filler items were used in both the direct and the indirect versions of each problem.

Table 3.1 presents the average ranks (R) of the conjunction $R(A\&B)$ and of its less representative constituents $R(B)$, relative to the set of five filler items. The percentage of violations of the conjunction rule in the direct test is denoted by V . The results can be summarized as follows: (a) the conjunction is ranked higher than its less likely constituents in all 12 comparisons, (b) there is no consistent difference between the ranks of the alternatives in the direct and indirect tests, (c) the overall incidence of violations of the conjunction rule in direct tests is 88%, which virtually coincides with the incidence of the corresponding pattern in judgments of representativeness, and (d) there is no effect of statistical sophistication in either indirect or direct tests.

Table 3.1 Tests of the conjunction rule in likelihood rankings

| Subjects | Problem | Direct test | | | | Indirect test | | |
|---------------|---------|-------------|-----------|-------|----|---------------|-------|---------|
| | | V | R (A & B) | R (B) | N | R (A & B) | R (B) | Total N |
| Naive | Bill | 92 | 2.5 | 4.5 | 94 | 2.3 | 4.5 | 88 |
| | Linda | 89 | 3.3 | 4.4 | 88 | 3.3 | 4.4 | 86 |
| Informed | Bill | 86 | 2.6 | 4.5 | 56 | 2.4 | 4.2 | 56 |
| | Linda | 90 | 3.0 | 4.3 | 53 | 2.9 | 3.9 | 55 |
| Sophisticated | Bill | 83 | 2.6 | 4.7 | 32 | 2.5 | 4.6 | 32 |
| | Linda | 85 | 3.2 | 4.3 | 32 | 3.1 | 4.3 | 32 |

Note. V = percentage of violations of the conjunction rule; R (A & B) and R (B) = mean rank assigned to A & B and to B, respectively; N = number of subjects in the direct test; Total N = total number of subjects in the indirect test, who were about equally divided between the two groups.

The violation of the conjunction rule in a direct comparison of B to A&B is called the *conjunction fallacy*. Violations inferred from between-subjects comparisons are called *conjunction errors*. Perhaps the most surprising aspect of Table 3.1 is the lack of any difference between indirect and direct tests. We had expected the conjunction to be judged more probable than the less likely of its constituents in an indirect test, in accord with the pattern observed in judgments of representativeness. However, we also expected that even naive respondents would notice the repetition of some attributes, alone and in conjunction with others, and that they would then apply the conjunction rule and rank the conjunction below its constituents. This expectation was violated, not only by statistically naive undergraduates but even by highly sophisticated respondents. In both direct and indirect tests, the subjects apparently ranked the outcomes by the degree to which Bill (or Linda) matched the respective stereotypes. The correlation between the mean ranks of probability and representativeness was .96 for Bill and .98 for Linda. Does the conjunction rule hold when the relation of inclusion is made highly transparent? The studies described in the next section abandon all subtlety in an effort to compel the subjects to detect and appreciate the inclusion relation between the target events.

Transparent Tests This section describes a series of increasingly desperate manipulations designed to induce subjects to obey the conjunction rule. We first presented the description of Linda to a group of 142 undergraduates at UBC and asked them to check which of two alternatives was more probable:

Linda is a bank teller. (T)

Linda is a bank teller and is active in the feminist movement. (T&F)

The order of alternatives was inverted for one half of the subjects, but this manipulation had no effect. Overall, 85% of respondents indicated

that T&F was more probable than T, in a flagrant violation of the conjunction rule.

Surprised by the finding, we searched for alternative interpretations of the subjects' responses. Perhaps the subjects found the question too trivial to be taken literally and consequently interpreted the inclusive statement T as T¬-F; that is, "Linda is a bank teller and is *not* a feminist." In such a reading, of course, the observed judgments would not violate the conjunction rule. To test this interpretation, we asked a new group of subjects ($N = 119$) to assess the probability of T and of T&F on a 9-point scale ranging from 1 (extremely unlikely) to 9 (extremely likely). Because it is sensible to rate probabilities even when one of the events includes the other, there was no reason for respondents to interpret T as T¬-F. The pattern of responses obtained with the new version was the same as before. The mean ratings of probability were 3.5 for T and 5.6 for T&F, and 82% of subjects assigned a higher rating to T&F than they did to T.

Although subjects do not spontaneously apply the conjunction rule, perhaps they can recognize its validity. We presented another group of UBC undergraduates with the description of Linda followed by the two statements, T and T&F, and asked them to indicate which of the following two arguments they found more convincing.

Argument 1. Linda is more likely to be a bank teller than she is to be a feminist bank teller, because every feminist bank teller is a bank teller, but some women bank tellers are not feminists, and Linda could be one of them.

Argument 2. Linda is more likely to be a feminist bank teller than she is likely to be a bank teller, because she resembles an active feminist more than she resembles a bank teller.

The majority of subjects (65%, $n = 58$) chose the invalid resemblance argument (Argument 2) over the valid extensional argument (Argument 1). Thus, a deliberate attempt to induce a reflective attitude did not eliminate the appeal of the representativeness heuristic.

We made a further effort to clarify the inclusive nature of the event T by representing it as a disjunction. (Note that the conjunction rule can also be expressed as a disjunction rule $P(A \text{ or } B) \geq P(B)$). The description of Linda was used again, with a 9-point rating scale for judgments of probability, but the statement T was replaced by

Linda is a bank teller whether or not she is active in the feminist movement. (T*)

This formulation emphasizes the inclusion of T&F in T. Despite the transparent relation between the statements, the mean ratings of likelihood were 5.1 for T&F and 3.8 for T* ($p < .01$, by t test). Furthermore, 75% of the subjects ($n = 75$) committed the conjunction fallacy by rating T&F higher than T*, and only 16% gave a lower rating to T&F than to T*.

The violations of the conjunction rule in direct comparisons of T&F to T* are remarkable because the extension of "Linda is a bank teller whether or not she is active in the feminist movement" clearly includes the extension of "Linda is a bank teller and is active in the feminist movement." Many subjects evidently failed to draw extensional inferences from the phrase "whether or not," which may have been taken to indicate a weak disposition. This interpretation was supported by a between-subjects comparison, in which different subjects evaluated T, T*, and T&F on a 9-point scale after evaluating the common filler statement, "Linda is a psychiatric social worker." The average ratings were 3.3 for T, 3.9 for T*, and 4.5 for T&F, with each mean significantly different from both others. The statements T and T* are of course extensionally equivalent, but they are assigned different probabilities. Because feminism fits Linda, the mere mention of this attribute makes T* more likely than T, and a definite commitment to it makes the probability of T&F even higher!

Modest success in loosening the grip of the conjunction fallacy was achieved by asking subjects to choose whether to bet on T or on T&F. The subjects were given Linda's description, with the following instruction:

If you could win \$10 by betting on an event, which of the following would you choose to bet on? (Check one)

The percentage of violations of the conjunction rule in this task was "only" 56% ($n = 60$), much too high for comfort but substantially lower than the typical value for comparisons of the two events in terms of probability. We conjecture that the betting context draws attention to the conditions in which one bet pays off whereas the other does not, allowing some subjects to discover that a bet on T dominates a bet on T&F.

The respondents in the studies described in this section were statistically naive undergraduates at UBC. Does statistical education eradicate the fallacy? To answer this question, 64 graduate students of social sciences at the University of California, Berkeley, and at Stanford University, all with credit for several statistics courses, were given the rating-scale version of the direct test of the conjunction rule for the Linda problem. For the first time in this series of studies, the mean rating for T&F (3.5) was lower than the rating assigned to T (3.8), and only 36% of respondents committed the fallacy. Thus, statistical sophistication produced a majority who conformed to the conjunction rule in a transparent test, although the incidence of violations was fairly high even in this group of intelligent and sophisticated respondents.

Elsewhere (Kahneman and Tversky, 1982), we distinguished between positive and negative accounts of judgments and preferences that violate normative rules. A positive account focuses on the factors that produce a particular response; a negative account seeks to explain why

the correct response was not made. The positive analysis of the Bill and Linda problems invokes the representativeness heuristic. The stubborn persistence of the conjunction fallacy in highly transparent problems, however, lends special interest to the characteristic question of a negative analysis: Why do intelligent and reasonably well-educated people fail to recognize the applicability of the conjunction rule in transparent problems? Postexperimental interviews and class discussions with many subjects shed some light on this question. Naive as well as sophisticated subjects generally noticed the nesting of the target events in the direct-transparent test, but the naive, unlike the sophisticated, did not appreciate its significance for probability assessment. On the other hand, most naive subjects did not attempt to defend their responses. As one subject said after acknowledging the validity of the conjunction rule, "I thought you only asked for my opinion."

The interviews and the results of the direct transparent tests indicate that naive subjects do not spontaneously treat the conjunction rule as decisive. Their attitude is reminiscent of children's responses in a Piagetian experiment. The child in the preconservation stage is not altogether blind to arguments based on conservation of volume and typically expects quantity to be conserved (Bruner 1966). What the child fails to see is that the conservation argument is decisive and should overrule the perceptual impression that the tall container holds more water than the short one. Similarly, naive subjects generally endorse the conjunction rule in the abstract, but their application of this rule to the Linda problem is blocked by the compelling impression that T&F is more representative of her than T is. In this context, the adult subjects reason as if they had not reached the stage of formal operations. A full understanding of a principle of physics, logic, or statistics requires knowledge of the conditions under which it prevails over conflicting arguments, such as the height of the liquid in a container or the representativeness of an outcome. The recognition of the decisive nature of rules distinguishes different developmental stages in studies of conservation; it also distinguishes different levels of statistical sophistication in the present series of studies. . . .

Cognitive Illusions Our studies of inductive reasoning have focused on systematic errors because they are diagnostic of the heuristics that generally govern judgment and inference. In the words of Helmholtz (1881/1903), "It is just those cases that are not in accordance with reality which are particularly instructive for discovering the laws of the processes by which normal perception originates." The focus on bias and illusion is a research strategy that exploits human error, although it neither assumes nor entails that people are perceptually or cognitively inept. Helmholtz's position implies that perception is not usefully analyzed into a normal process that produces accurate percepts and a distorting process that produces errors and illusions. In cognition, as

in perception, the same mechanisms produce both valid and invalid judgments. Indeed, the evidence does not seem to support a “truth plus error” model, which assumes a coherent system of beliefs that is perturbed by various sources of distortion and error. Hence, we do not share Dennis Lindley’s optimistic opinion that “inside every incoherent person there is a coherent one trying to get out” (Lindley, personal communication 1980), and we suspect that incoherence is more than skin deep (Tversky and Kahneman, 1981).

It is instructive to compare a structure of beliefs about a domain, (e.g., the political future of Central America) to the perception of a scene (e.g., the view of Yosemite Valley from Glacier Point). We have argued that intuitive judgments of all relevant marginal, conjunctive, and conditional probabilities are not likely to be coherent, that is, to satisfy the constraints of probability theory. Similarly, estimates of distances and angles in the scene are unlikely to satisfy the laws of geometry. For example, there may be pairs of political events for which $P(A)$ is judged greater than $P(B)$ but $P(A/B)$ is judged less than $P(B/A)$ —see Tversky and Kahneman (1980). Analogously, the scene may contain a triangle ABC for which the A angle appears greater than the B angle, although the BC distance appears to be smaller than the AC distance.

The violations of the qualitative laws of geometry and probability in judgments of distance and likelihood have significant implications for the interpretation and use of these judgments. Incoherence sharply restricts the inferences that can be drawn from subjective estimates. The judged ordering of the sides of a triangle cannot be inferred from the judged ordering of its angles, and the ordering of marginal probabilities cannot be deduced from the ordering of the respective conditionals. The results of the present study show that it is even unsafe to assume that $P(B)$ is bounded by $P(A\&B)$. Furthermore, a system of judgments that does not obey the conjunction rule cannot be expected to obey more complicated principles that presuppose this rule, such as Bayesian updating, external calibration, and the maximization of expected utility. The presence of bias and incoherence does not diminish the normative force of these principles, but it reduces their usefulness as descriptions of behavior and hinders their prescriptive applications. Indeed, the elicitation of unbiased judgments and the reconciliation of incoherent assessments pose serious problems that presently have no satisfactory solution (Lindley et al. 1979; Shafer and Tversky 1983).

The issue of coherence has loomed larger in the study of preference and belief than in the study of perception. Judgments of distance and angle can readily be compared to objective reality and can be replaced by objective measurements when accuracy matters. In contrast, objective measurements of probability are often unavailable, and most significant choices under risk require an intuitive evaluation of probability. In the absence of an objective criterion of validity, the normative theory of judgment under uncertainty has treated the coherence of belief as

the touchstone of human rationality. Coherence has also been assumed in many descriptive analyses in psychology, economics, and other social sciences. This assumption is attractive because the strong normative appeal of the laws of probability makes violations appear implausible. Our studies of the conjunction rule show that normatively inspired theories that assume coherence are descriptively inadequate, whereas psychological analyses that ignore the appeal of normative rules are, at best, incomplete. A comprehensive account of human judgment must reflect the tension between compelling logical rules and seductive non-extensional intuitions.

Note

Research on the second half of this chapter was supported by Grant NR 197-058 from the U.S. Office of Naval Research.

References

- Bruner, J. S. (1966). On the conservation of liquids. In J. S. Bruner, R. R. Oliver, and P. M. Greenfield et al., eds., *Studies in cognitive growth*. New York: Wiley.
- Chapman, L. J., and J. P. Chapman (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology* 73, 193–204.
- Chapman, L. J. and J. P. Chapman (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology* 74, 271–280.
- Cohen, L. J. (1977). *The probable and the provable*. Oxford: Clarendon Press.
- Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*. 38, 325–339.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz, ed., *Formal representation of human judgment*. New York: Wiley.
- Galbraith, R. C., and B. J. Underwood (1973). Perceived frequency of concrete and abstract words. *Memory and Cognition* 1, 56–60.
- Gati, I., and A. Tversky (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human perception and Performance* 8, 325–340.
- Helmholtz, H. von (1881/1903). *Popular lectures on scientific subjects*, trans. E. Atkinson. New York: Green.
- Jennings, D., T. Amabile, and L. Ross (1982). Informal covariation assessment. In D. Kahneman, P. Slovic, and A. Tversky, eds., *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kahneman, D., P. Slovic, and A. Tversky, eds. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

- Kahneman, D., and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3, 430–454.
- Kahneman, D., and A. Tversky (1973). On the psychology of prediction. *Psychological Review* 80, 237–251.
- Kahneman, D., and A. Tversky (1982). On the study of statistical intuitions. *Cognition* 11, 123–141.
- Kyburg, H. E. (1983). Rational belief. *Behavioral and Brain Sciences* 6, 231–273.
- Lindley, D. V., A. Tversky, and R. V. Brown (1979). On the reconciliation of probability assessments. *Journal of the Royal Statistical Society* 142, 146–180.
- Mervis, C. B., and E. Rosch (1981). Categorization of natural objects. *Annual Review of Psychology* 32, 89–115.
- Nisbett, R. E., and L. Ross (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Rosch, E. (1978). Principles of categorization. In E. Rosch and B. B. Lloyd, eds., *Cognition and categorization*. Hillsdale, NJ: Erlbaum.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.
- Shafer, G., and A. Tversky (1983). Weighing evidence: The design and comparison of probability through experiments. Unpublished manuscript, Stanford University, Stanford.
- Slovic, P. and S. Lichtenstein (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance* 6, 649–744.
- Smith, E. E., and D. L. Medin (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Suppes, P. (1975). Approximate probability and expectation of gambles. *Erkenntnis* 9, 153–161.
- Tversky, A. (1977). Features of similarity. *Psychological Review* 84, 327–352.
- Tversky, A., and D. Kahneman (1971). Belief in the “law of small numbers.” *Psychological Bulletin* 76, 105–110.
- Tversky, A., and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5, 207–232.
- Tversky, A., and D. Kahneman (1980). Causal schemas in judgments under uncertainty. In M. Fishbein, ed., *Progress in social psychology*. Hillsdale, NJ: Erlbaum.
- Tversky, A., and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458.

Tversky, A., and D. Kahneman (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, and A. Tversky, eds., *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

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