Judgment under Uncertainty: Heuristics and Biases

Biases in judgments reveal some heuristics of thinking under uncertainty.

Amos Tversky and Daniel Kahneman

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 task 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. Such biases are also found in the intuitive judgment of probability. This article describes three heuristics 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 (1). 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 experiment where prior probabilities were manipulated (1).

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 essen-
tially the same probability judgments. Apparently, subjects evaluated the like-
lihood that a particular description be-
longed to an engineer rather than to a
lawyer by the degree to which this de-
scription was representative of the
two stereotypes, with little or no regard
for the prior probabilities of the cate-
gories.

The subjects used prior probabilities
correctly when they had no other infor-
mation. 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 uninfor-
mative. The responses to the follow-

This description was intended to con-
vey no information relevant to the ques-
tion 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 engi-
neer 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 evi-
dence and when given worthless evi-
dence. When no specific evidence is
given, prior probabilities are properly
utilized; when worthless evidence is
given, prior probabilities are ignored
(1).

Insensitivity to sample size. To evalu-
ate the probability of obtaining a par-
ticular 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 centi-
meters), by the similarity of this result
to the corresponding parameter (that
is, to the average height in the popula-
tion of men). The similarity of a sam-
ple 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 sta-
tistic 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 obtain-
ing an average height greater than 6
feet was assigned the same value for
samples of 1000, 100, and 10 men (2).
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 hospi-
tals. 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. Some-
times 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 num-
er 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 sta-
tistic and are therefore equally repre-
sentative 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 hospi-
tal 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 pos-
terior probability, that is, of the prob-
ability that a sample has been drawn
from one population rather than from
another. Consider the following ex-
ample:

Imagine an urn filled with balls, of
which 5/6 are of one color and 1/6 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 5/6 red balls and 1/6
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 5:1 sample
and 16 to 1 for the 12:8 sample, as-
suming equal prior probabilities. How-
ever, most people feel that the first
sample provides much stronger evidence
for the hypothesis that the urn is pre-
dominantly 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 (2). In addi-
tion, intuitive estimates of posterior
odds are far less extreme than the cor-
rect values. The underestimation of
the impact of evidence has been observed
repeatedly in problems of this type (2, 4).
It has been labeled "conservatism."

Misconceptions of chance. People ex-
pect 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-H to be more likely
than the sequence H-H-H-T-T, which
does not appear random, and also more likely than the sequence
H-H-H-T-H, which does not represent
the fairness of the coin (2). 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, how-
ever, deviates systematically from chance
expectation: it contains too many al-
ternations and too few runs. Another
consequence of the belief in local rep-
resentativeness is the well-known gam-
blers' fallacy. After observing a long
run of red on the roulette wheel, for
example, most people erroneously be-
lieve 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 devi-
ation in one direction induces a devia-
tion 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 (5) reveals a
lingering belief in what may be called
the "law of small numbers," according
to which even small samples are highly

27 SEPTEMBER 1974

1125
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.

Inconsistency 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 (1). 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 (1).

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 mean 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 (1). 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 (1). 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 interactions, 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 recalled 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 (6).

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 (6).

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 (7) 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 ≤ k ≤ 8. How many different committees of k members can be formed? The correct answer to this problem is given by the binomial coefficient (10 choose 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 fewer 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 (6). 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 imagining 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 (6) 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.

**Adjustment and Anchoring**

In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient (4).

That is, different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring.

**Insufficient adjustment.** In a demonstration of the anchoring effect, subjects were asked to estimate various quantities, stated in percentages (for example, the percentage of African countries in the United Nations). For each quantity, a number between 0 and 100 was determined by spinning a wheel of fortune in the subjects' presence. The subjects were instructed to indicate first whether that number was higher or lower than the value of the quantity, and then to estimate the value of the quantity by moving upward or downward from the given number. Different groups were given different numbers for each quantity, and these arbitrary numbers had a marked effect on estimates. For example, the median estimates of the percentage of African countries in the United Nations were 25 and 45 for groups that received 10 and 65, respectively, as starting points. Payoffs for accuracy did not reduce the anchoring effect.

Anchoring occurs not only when the starting point is given to the subject, but also when the subject bases his estimate on the result of some incomplete computation. A study of intuitive numerical estimation illustrates this effect. Two groups of high school students estimated, within 5 seconds, a numerical expression that was written on the blackboard. One group estimated the product

\[ 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 \]

while another group estimated the product

\[ 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \]

To rapidly answer such questions, people may perform a few steps of computation and estimate the product by extrapolation or adjustment. Because adjustments are typically insufficient, this procedure should lead to underestimation. Furthermore, because the result of the first few steps of multiplication (performed from left to right) is higher in the descending sequence than in the ascending sequence, the former expression should be judged larger than the latter. Both predictions were confirmed. The median estimate for the ascending sequence was 512, while the median estimate for the descending sequence was 2,250. The correct answer is 40,320.

**Biases in the evaluation of conjunctive and disjunctive events.** In a recent
study by Bar-Hillel (9) subjects were given the opportunity to bet on one of two events. Three types of events were used: (i) simple events, such as drawing a red marble from a bag containing 50 percent red marbles and 50 percent white marbles; (ii) conjunctive events, such as drawing a red marble seven times in succession, with replacement, from a bag containing 90 percent red marbles and 10 percent white marbles; and (iii) disjunctive events, such as drawing a red marble at least once in seven successive tries, with replacement, from a bag containing 10 percent red marbles and 90 percent white marbles. In this problem, a significant majority of subjects preferred to bet on the conjunctive event (the probability of which is .49) rather than on the simple event (the probability of which is .50). Subjects also preferred to bet on the simple event rather than on the disjunctive event, which has a probability of .52. Thus, most subjects bet on the less likely event in both comparisons. This pattern of choices illustrates a general finding. Studies of choice among gambles and of judgments of probability indicate that people tend to overestimate the probability of conjunctive events (10) and to underestimate the probability of disjunctive events. These biases are readily explained as effects of anchoring. The stated probability of the elementary event (success at any one stage) provides a natural starting point for the estimation of the probabilities of both conjunctive and disjunctive events. Since adjustment from the starting point is typically insufficient, the final estimates remain too close to the probabilities of the elementary events in both cases. Note that the overall probability of a conjunctive event is lower than the probability of each elementary event, whereas the overall probability of a disjunctive event is higher than the probability of each elementary event. As a consequence of anchoring, the overall probability will be overestimated in conjunctive problems and underestimated in disjunctive problems.

Biases in the evaluation of compound events are particularly significant in the context of planning. The successful completion of an undertaking, such as the development of a new product, typically has a conjunctive character: for the undertaking to succeed, each of a series of events must occur. Even when each of these events is very likely, the overall probability of success can be quite low if the number of events is large. The general tendency to overestimate the probability of conjunctive events leads to unwarranted optimism in the evaluation of the likelihood that a plan will succeed or that a project will be completed on time. Conversely, disjunctive structures are typically encountered in the evaluation of risks. A complex system, such as a nuclear reactor or a human body, will malfunction if any of its essential components fails. Even when the likelihood of failure in each component is slight, the probability of an overall failure can be high if many components are involved. Because of anchoring, people will tend to underestimate the probabilities of failure in complex systems. Thus, the direction of the anchoring bias can sometimes be inferred from the structure of the event. The chain-like structure of conjunctions leads to overestimation, the funnel-like structure of disjunctions leads to underestimation.

Anchoring in the assessment of subjective probability distributions. In decision analysis, experts are often required to express their beliefs about a quantity, such as the value of the Dow-Jones average on a particular day, in the form of a probability distribution. Such a distribution is usually constructed by asking the person to select values of the quantity that correspond to specified percentiles of his subjective probability distribution. For example, the judge may be asked to select a number, X_{90}, such that his subjective probability that this number will be higher than the value of the Dow-Jones average is .90. That is, he should select the value X_{90} so that he is just willing to accept 9 to 1 odds that the Dow-Jones average will not exceed it. A subjective probability distribution for the value of the Dow-Jones average can be constructed from several judgments corresponding to different percentiles.

By collecting subjective probability distributions for many different quantities, it is possible to test the judge for proper calibration. A judge is properly (or externally) calibrated in a set of problems if exactly 11 percent of the true values of the assessed quantities falls below his stated values of Xn. For example, the true values should fall below X_{01} for 1 percent of the quantities and above X_{99} for 1 percent of the quantities. Thus, the true values should fall in the confidence interval between X_{01} and X_{99} on 98 percent of the problems. Several investigators (11) have obtained probability distributions for many quantities from a large number of judges. These distributions indicated large and systematic departures from proper calibration. In most studies, the actual values of the assessed quantities are either smaller than X_{01} or greater than X_{99} for about 50 percent of the problems. That is, the subjects state overly narrow confidence intervals which reflect more certainty than is justified by their knowledge about the assessed quantities. This bias is common to naive and to sophisticated subjects, and it is not eliminated by introducing proper scoring rules, which provide incentives for external calibration. This effect is attributable, in part at least, to anchoring.

To select X_{90} for the value of the Dow-Jones average, for example, it is natural to begin by thinking about one's best estimate of the Dow-Jones and to adjust this value upward. If this adjustment—like most others—is insufficient, then X_{90} will not be sufficiently extreme. A similar anchoring effect will occur in the selection of X_{10}, which is presumably obtained by adjusting one's best estimate downward. Consequently, the confidence interval between X_{0} and X_{90} will be too narrow, and the assessed probability distribution will be too tight. In support of this interpretation it can be shown that subjective probabilities are systematically altered by a procedure in which one's best estimate does not serve as an anchor.

Subjective probability distributions for a given quantity (the Dow-Jones average) can be obtained in two different ways: (i) by asking the subject to select values of the Dow-Jones that correspond to specified percentiles of his probability distribution and (ii) by asking the subject to assess the probabilities that the true value of the Dow-Jones will exceed some specified values. The two procedures are formally equivalent and should yield identical distributions. However, they suggest different modes of adjustment from different anchors. In procedure (i), the natural starting point is one's best estimate of the quantity. In procedure (ii), on the other hand, the subject may be anchored on the value stated in the question. Alternatively, he may be anchored on even odds, or 50-50 chances, which is a natural starting point in the estimation of likelihood. In either case, procedure (ii) should yield less extreme odds than procedure (i).

To contrast the two procedures, a set of 24 quantities (such as the air dis-
It is not surprising that useful heuristics such as representativeness and availability are retained, even though they occasionally lead to errors in prediction or estimation. What is perhaps surprising is the failure of people to infer from lifelong experience such fundamental statistical rules as regression toward the mean, or the effect of sample size on sampling variability. Although everyone is exposed, in the normal course of life, to numerous examples from which these rules could have been induced, very few people discover the principles of sampling and regression on their own. Statistical principles are not learned from everyday experience because the relevant instances are not coded appropriately. For example, people do not discover that successive lines in a text differ more in average word length than do successive pages, because they simply do not attend to the average word length of individual lines or pages. Thus, people do not learn the relation between sample size and sampling variability, although the data for such learning are abundant.

The lack of an appropriate code also explains why people usually do not detect the biases in their judgments of probability. A person could conceivably learn whether his judgments are externally calibrated by keeping a tally of the proportion of events that actually occur among those to which he assigns the same probability. However, it is not natural to group events by their judged probability. In the absence of such grouping it is impossible for an individual to discover, for example, that only 50 percent of the predictions to which he has assigned a probability of .9 higher actually came true.

The empirical analysis of cognitive biases has implications for the theoretical and applied role of judged probabilities. Modern decision theory (12, 13) regards subjective probability as the quantified opinion of an idealized person. Specifically, the subjective probability of a given event is defined by the set of bets about this event that such a person is willing to accept. An internally consistent, or coherent, subjective probability measure can be derived for an individual if his choices among bets satisfy certain principles, that is, the axioms of the theory. The derived probability is subjective in the sense that different individuals are allowed to have different probabilities for the same event.

The major contribution of this approach is that it provides a rigorous subjective interpretation of probability that is applicable to unique events and is embedded in a general theory of rational decision.

It should perhaps be noted that, while subjective probabilities can sometimes be inferred from preferences among bets, they are normally not formed in this fashion. A person bets on team A rather than on team B because he believes that team A is more likely to win; he does not infer this belief from his betting preferences. Thus, in reality, subjective probabilities determine preferences among bets and are not derived from them, as in the axiomatic theory of rational decision (12).

The inherently subjective nature of probability has led many students to the belief that coherence, or internal consistency, is the only valid criterion by which judged probabilities should be evaluated. From the standpoint of the formal theory of subjective probability, any set of internally consistent probabilities is as good as any other. This criterion is not entirely satisfactory, because an internally consistent set of subjective probabilities can be incompatible with other beliefs held by the individual. Consider a person whose subjective probabilities for all possible outcomes of a coin-tossing game reflect the gambler's fallacy. That is, his estimate of the probability of tails on a particular toss increases with the number of consecutive heads that preceded that toss. The judgments of such a person could be internally consistent and therefore acceptable as adequate subjective probabilities according to the criterion of the formal theory. These probabilities, however, are incompatible with the generally held belief that a coin has no memory and is therefore incapable of generating sequential dependencies. For judged probabilities to be considered adequate, or rational, internal consistency is not enough. The judgments must be compatible with the entire web of beliefs held by the individual. Unfortunately, there can be no simple formal procedure for assessing the compatibility of a set of probability judgments with the judge's total system of beliefs. The rational judge will nevertheless strive for compatibility, even though internal consistency is more easily achieved and assessed. In particular, he will attempt to make his probability judgments compatible with his knowledge about the subject matter, the laws of probability, and his own judgmental heuristics and biases.
Rural Health Care in Mexico?

Present educational and administrative structures must be changed in order to improve health care in rural areas.

Luis Cañedo

The present health care structure in Mexico focuses attention on the urban population, leaving the rural communities practically unattended. There are two main factors contributing to this situation. One is the lack of coordination among the different institutions responsible for the health of the community and among the educational institutions. The other is the lack of information concerning the nature of the problems in rural areas. In an attempt to provide a solution to these problems, a program has been designed that takes into consideration the environmental conditions, malnutrition, poverty, and negative cultural factors that are responsible for the high incidences of certain diseases among rural populations. It is based on the development of a national information system for the collection and dissemination of information related to general, as well as rural, health care, that will provide the basis for a national health care system, and depends on the establishment of a training program for professionals in community medicine.

The continental and insular area of Mexico, including interior waters, is 2,022,058 square kilometers (1, 2). In 1970 the population of Mexico was 48,377,363, of which 24,055,305 persons (49.7 percent) were under 15 years of age. The Indian population made up 7.9 percent of the total (2, 3). As indicated in Table 1, 42.3 percent of the total population live in communities of less than 2,500 inhabitants, and in such communities public services as well as means of communication are very scarce or nonexistent. A large percentage (39.5 percent) of the economically active population is engaged in agriculture (4).

The country's population growth rate is high, 3.5 percent annually, and it seems to depend on income, being higher among the 50 percent of the population earning less than 675 pesos ($50) per family per month (5). The majority of this population lives in the rural areas. The most frequent causes of mortality in rural areas are malnutrition, infectious and parasitic diseases (6, 7), pregnancy complications, and accidents (2). In 1970 there were 34,107 doctors in Mexico (2). The ratio of inhabitants to doctors, which is 1423.7, is not a representative index of the actual distribution of resources because there is a great scarcity of health professionals in rural areas and a high concentration in urban areas (Fig. 1) (7, 8).

In order to improve health at a national level, this situation must be changed. The errors made in previous attempts to improve health care must be avoided, and use must be made of the available manpower and resources of modern science to produce feasible answers at the community level. Although the main objective of a specialist in community medicine is to control disease, such control cannot be achieved unless action is taken against the underlying causes of disease; it has already been observed that partial solutions are inefficient (9). As a background to this new program that has been designed to provide health care in rural communities, I shall first give a summary of the previous attempts that have been made to provide such care, describing the various medical institutions and other organizations that are responsible for the training of medical personnel and for constructing the facilities required for health care.

The author is an investigator in the department of molecular biology of the Instituto de Investigaciones Biomedicas, Universidad Nacional Autonoma de Mexico, Ciudad Universitaria, Mexico 04, D.F. This article is adapted from a paper presented at the meeting on Science and Man in the Americas, jointly organized by the Consejo Nacional de Ciencia y Tecnologia de Mexico and the American Association for the Advancement of Science and held in Mexico City, 20 June to 4 July 1973.

27 SEPTEMBER 1974