1

Outline of heuristics and biases

Summary

Probabilities differ from numbers in that they have a limited range of from 0.0 through 1.0. Probability theory and statistics were developed between World Wars I and II in order to model events in the outside world. More recently, the models have been used to describe the way people think, or ought to think. Both frequentists and some subjectivists argue that the average degree of confidence in single events is not comparable to the average judged frequency of events in a long run. The discrepancies between the 2 measures can be ascribed to biases in the different ways of responding.

Tversky and Kahneman, and their likeminded colleagues, describe a number of what can be called heuristic or complex biases that influence the way people deal with probabilities: apparent overconfidence, hindsight bias, the small sample fallacy, the conjunction fallacy, the regression fallacy, base rate neglect, the availability and simulation fallacies, the anchoring and adjustment biases, the expected utility fallacy, and bias by frames. Each complex bias can be described by a heuristic or rule of thumb that can be said to be used instead of the appropriate normative rule. Some version of the heuristic of representativeness is used most frequently.

While studying a complex bias, an investigator may introduce, without realizing it, a simple bias of the kind that is found also in quantifying the judgments of sensory psychophysics. The relevant simple biases are: the response contraction bias or regression effect, the sequential contraction bias, the stimulus range equalizing bias, the equal frequency bias, and the transfer bias.

Probabilities

Many people are not familiar with probabilities. Probabilities appear to be numbers, but they have a small fixed range of from 0.0 through 1.0, which distinguishes them from numbers. Probabilities can be added, averaged and multiplied, as can numbers, but they can also be combined by other methods described in Chapter 2. Their small fixed range of from 0.0 through 1.0 suggests a reference magnitude of .5 in the middle. The reference magnitude encourages the response contraction bias or regression effect, which is described later in the chapter.

Probabilities are not much used in everyday life. They are more likely to be stated as ratios like one chance in 1,000, or like the odds used in betting. A probability of .8 is described as odds of .8 to .2, or 4 to 1. Figure 1.1C shows a scale of odds with the odds plotted logarithmically. The full scale would extend from 1:1 on the left to infinity: 1 on the extreme right. The scale does not have a small fixed range like a scale of probability, and so lacks an obvious reference magnitude.

Percentages are proportions that have some of the characteristics of probabilities. Figure 1.1A shows that the range 0% through 100% can be used as a substitute for probabilities ranging from 0.0 through 1.0. But percentages can be negative, like a 3% discount. Also they can exceed 100%, as when an object is sold for more than double its original cost. An advantage of using percentages as substitutes for probabilities is that people are more likely to be familiar with percentages. But percentages can present difficulties to students who are not familiar with them.

Frequencies can be used as substitutes for both probabilities and percentages. Out of 100 items, a frequency of 50 corresponds to a probability of .5 or 50%. Most ordinary people find frequencies easier to estimate than either percentages or probabilities.

A brief history of probability in the behavioral sciences

The mathematics of probability has been available for the past 200 years. In 1774 Laplace (quoted by Gigerenzer and Murray, 1987, p. 147) proved Bayes' theorem, but it was the period between World Wars I and II that saw the increasing reliance on the use of probability theory and statistics. In the natural sciences 2 rather similar schools developed (after Gigerenzer, 1993). Both schools used the same statistical tests, such as the **t** and F tests, but they had different practical aims. The strict school of Neyman and Pearson grew out of the requirements for quality control in

Cambridge University Press
978-0-521-44368-5 - Behavioral Decisiontheory: A New Approach
E.C. Poultion
Excerpt
More information

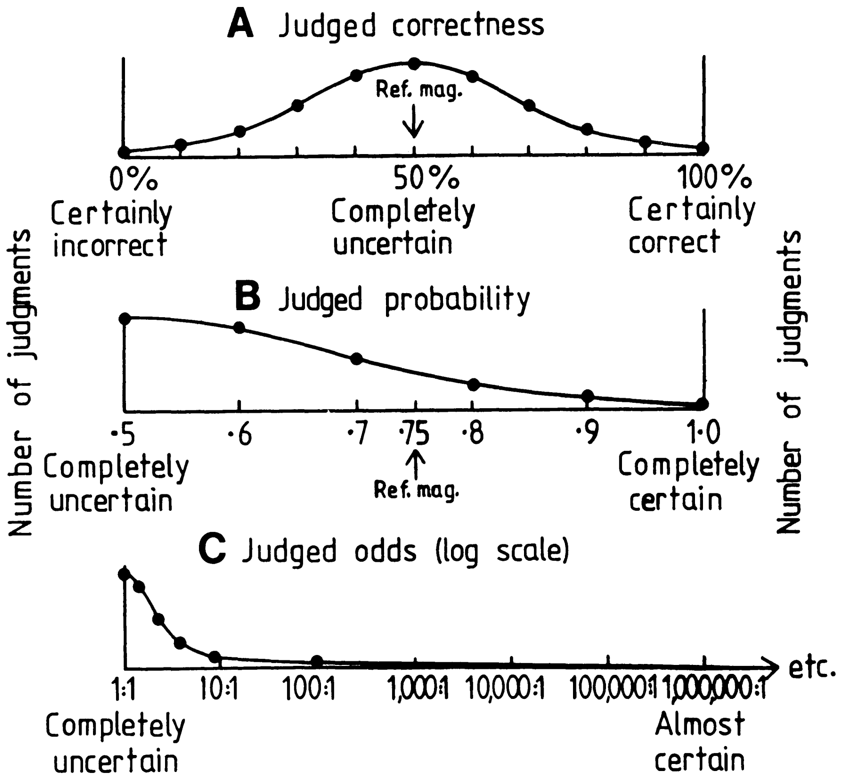

Figure 1.1. Rating scales for probability. Scale A is used in judging the percent probability of correctness of items that can be correct or incorrect. The theoretical function is a normal or Gaussian distribution with a mean of 50%, or a mean probability of 0.5. The arrow shows that 50% is usually the reference magnitude for the response contraction bias.

Scale B is an expanded version of the right half of Scale A. It is shown calibrated in units of probability, but it can be calibrated in percent probabilities like Scale A. Scale B is used when there is a choice between a pair of items. If the item chosen as correct is given a probability less than 0.5, the other item has a probability greater than .5 and so should be chosen instead. The arrow shows that here the reference magnitude for the response contraction bias is usually .75, which encourages overstimation.

Scale C is a logarithmic version of scale B, calibrated in odds. The scale has no upper limit. Completely certain judgments would be infinitely far to the right. The scale avoids the response contraction bias because it has no obvious reference magnitude. But under the influence of the equal frequency bias the logarithmic calibration marks can produce too many highly overconfident judgments.

manufacturing industry. The aim was to enable managers to decide the point at which rejected products, whether faulty or up to standard but incorrectly rejected, were sufficiently numerous to make it necessary to shut down the plant and overhaul it. The method characteristically involves repeated random sampling using previously specified values such as the proportion of incorrect rejections or Type 1 errors, statistical power, the number of samples, or previously specified hypotheses.

Fisher developed his less strict school primarily to deal with the results of the growing number of agricultural field trials. The aim was to enable investigators to check their calculations and conclusions. The probability values derived from **t** or F tests were to be evaluated in the light of the initial hypothesis: the probability of the data given the hypothesis, or p(D/H). As in astronomy, occasional data points could be rejected if they were too far removed from the mean.

Since World War II a third school, that of Bayesian statistics, has become more popular. This school characteristically deals with the inverse probability: the probability of the hypothesis given the data, or p(H/D). The investigators develop a number of hypotheses and compare them on how well they account for the data. The method was condemned by Fisher as being too subjective. The main application of Bayesian statistics discussed in this book is in combining 2 probabilities such as a base rate or prior probability, and a likelihood probability. When expressed as odds,

$$\text{Prior odds} \times \text{likelihood ratio} = \text{posterior odds} \qquad (1.1)$$

In these applications the meaning of probability changes from the philosophical degree of subjective certainty in the mind to the practical ratio of objective events counted in the outside world (Gigerenzer, Swijtink, Porter, Daston, Beatty and Krüger, 1989).

More recently probability theory and statistics have come to be used to provide models for the way people think—at first the way trained investigators think, but later the way ordinary untrained people think, or ought to think. Thus, when there are prior probabilities or base rates, it is assumed that people should estimate changes in probability using the Bayes method of Equation (1.1). Deviations from the Bayes model come to be explained using the vocabulary of the Bayes model, e.g., the miscalculation or misaggregation of probabilities (Gigerenzer and Murray, 1987, Chapter 5).

Investigators may use a double standard. In solving problems they may expect their respondents to use the Bayes method, but in evaluating the reliability of their respondents' solutions, they themselves may use the conventional mixture of Fisherian and Neyman–Pearson statistics like the **t** or F test (Gigerenzer, 1991a, p. 261). Gigerenzer, Hell and Blank (1988, pp. 516 and 521) use this double standard.

Alternative interpretations of probability

As Gigerenzer (1991a, p. 259) points out, there may be a choice of statistical measures to use, and no obvious reason why one should be used rather than another. For example, after choosing one of 2 answers, students could be asked to estimate the probability that their choice is correct. The students could be supplied with a scale of probabilities ranging from 0.0 through 1.0. The investigator obtains a number of confidence judgments and computes some kind of average probability. Other students could be asked simply to state, after making a series of choices, what proportion they judge to be correct. A theoretically perfectly calibrated person should provide a distribution of confidence judgments that corresponds exactly to the judged proportion of successes. But this leaves open the question of what measure of the distribution of confidence judgments to use. When the distribution is skewed, should the mean, the median, or the mode of the distribution be taken to correspond to the judged proportion of correct responses? Arguments could be presented to support any one of the alternatives. The alternative criteria for matching distributions of confidence judgments to judged relative frequencies are among a number of examples where statistics do not provide a single unambiguous answer.

The disagreements have their counterparts in theories of probability. Gigerenzer (1991a, p. 261) describes 2 theoretical approaches to probability that avoid the problem of comparing a distribution of confidence judgments with the judged frequency of correct answers in a long run. According to frequentists, probability theory does not apply to confidence in single events because probability theory is about frequencies in a long run. Frequentists believe that the term probability has no meaning when it refers to single events. On this view, a discrepancy between the degree of confidence in single events and the judged relative frequency in a long run should not be called a bias in probabilistic reasoning, because it involves comparing 2 measures that are not comparable.

A contrary point of view is held by some subjectivists. To them probability is about single events, but rationality is identified with the internal consistency of probability judgments. On this view, in whatever way an individual evaluates the probability of a particular event, no experience can prove him or her right or wrong because there is no criterion here to distinguish between right and wrong. Thus, for different reasons, both interpretations suggest that the average degree of confidence in the correctness of single answers is not comparable to the estimated relative frequency of correct answers in a long run.

Few practical investigators who study confidence judgments appear to pay attention to these conflicting interpretations and their implications. It is not clear whether this should be blamed upon the arbitrary nature of the theoretical interpretations, or upon the failure of communication between the theorists and the investigators. The course of action followed in this book is to note the theoretical distinction between confidence in the correctness of single judgments and the proportion of judged right and wrong answers in a long run. It is assumed that the mismatch between the different measures is likely to be a small second order effect when it is compared with the sizes of the biases that are introduced by the different methods of responding.

Heuristic or complex biases in dealing with probabilities

Subjective probabilities show 2 kinds of bias. First, most ordinary people know little or nothing about probability theory and statistics. Thus, their estimates of probability are likely to be biased in a number of ways. This first kind of bias is here called heuristic or complex. The second kind of bias occurs more widely, in dealing with both probabilities and sensory magnitudes (Poulton, 1979, 1989). It may be introduced unintentionally by the investigators themselves. It is here called simple, to distinguish if from the heuristic or complex biases. In different places the name bias is used to describe both the effect and the psychological mechanism. Thus, a bias can be said both to specify a kind of biased judgment and to produce the judgment.

Table 1.1 lists the heuristic or complex biases in estimating probabilities that are studied by Tversky and Kahneman, and their like minded colleagues. Each heuristic or complex bias can be described as the failure to use an appropriate normative rule. For the 6 biases marked with an asterisk, Tversky and Kahneman describe the heuristics or rules of thumb that they believe to be used instead of using the normative rules. The table shows the remaining complex biases classified in the same framework of heuristic rules of thumb and normative rules, in keeping with Tversky and Kahneman's distinctions.

Table 1.1. *Heuristic or complex biases in dealing with probabilities*

Chapter	Complex bias	Normative rule	Heuristic bias
3.	Apparent overconfidence	Use objective probability	Use probability of related knowledge
4.	Hindsight bias	Avoid using hindsight knowledge	Use hindsight knowledge
5.	Small sample fallacy*	Small samples are not as representative as are large samples	Small and large samples should be equally representative
6.	Conjunction fallacy*	A conjunction is less probable than either component	A conjunction is more representative than is its less probable component
7.	Regression fallacy*	Future scores regress towards the average	Future scores should be maximally representative of past scores, and so should not regress
8.	Base rate neglect*	Combine 2 independent probabilities of the same event using the Bayes method	Ignore or give less weight to the prior probability or base rate
9.	Availability and simulation fallacies*	Use objective measures of frequency or probability	Judge frequency or probability from availability in memory or ease of simulation
10.	Anchoring and adjustment biases*	Avoid using anchors	Make use of anchors
11.	Expected utility fallacy	Choose largest expected gain or smallest expected loss	Choose certain gains and avoid certain losses, unless one or both probabilities are very low
12.	Bias by frames	Avoid using frames	Use frames

*Heuristic described by Tversky and Kahneman

Cambridge University Press
978-0-521-44368-5 - Behavioral Decisiontheory: A New Approach
E.C. Poultion
Excerpt
More information

Apparent overconfidence

Suppose people are set questions and are asked to estimate the probability that their answers are correct. If they have to use a response scale that is biased towards overestimation, they are likely to appear overconfident (Fischhoff and MacGregor, 1982; Lichtenstein and Fischhoff, 1977). People can also be said to show overconfidence in setting uncertainty bounds on unknown quantities using the fractile method of Chapter 3. The uncertainty bounds are set too close together. This is conventionally described as overconfidence. But it can be produced by a simple bias of quantification, the response contraction bias or regression effect, which is described in the last part of this chapter. A possible reason for overconfidence in answering knowledge questions is that people tend to use the heuristic of estimating the probability of their knowledge of the area (Glenberg, Sanocki, Epstein and Morris, 1987). They do not follow the normative rule and use their detailed knowledge of the question asked.

Hindsight bias

In predicting the uncertain outcome of a future event, forecasters are likely to compare the present circumstances with the circumstances surrounding the outcomes of previous events of a similar kind. They may then judge the relative importance of the influences that they believe will combine to determine the outcome, and make the prediction that is most compatible with this knowledge.

After the event, hindsight is likely to change the judged relative importance of the influences in directions that make them more compatible with the known outcome. This restructuring makes the outcome appear more probable with hindsight than it does in forecasting (Fischhoff, 1975b, 1977). It encourages the heuristic of using hindsight knowledge, instead of following the normative rule and avoiding it.

Small sample fallacy

The small sample fallacy occurs when a population is heterogeneous. Tversky and Kahneman's (1971, p. 106) heuristic bias is based on the belief that samples of all sizes should be equally representative of the parent population. The normative rule is that small heterogeneous samples are not as representative as are large samples. Representativeness can be defined along 3 different dimensions, one dimension for each of 3 different versions of the fallacy.

What can be called the small sample fallacy for size or reliability is the belief that samples of all sizes should be equally reliable. What can be called the small sample fallacy for distributions is the belief that small and large sample distributions should be equally regular, but not too regular. The gamblers' fallacy is another small sample fallacy for distributions. An example is the belief that after a run of coin tosses coming down 'heads', a 'tails' is more likely. The corresponding dimension of representativeness could be said to be the belief that samples of all sizes are equally self-correcting.

The small sample fallacy may be committed in investigations of other complex biases. This can happen if an investigator fails to take account of sample size or variability when comparing or combining probabilities that are expressed as percentages or frequencies. Whenever probabilities are expressed as percentages or frequencies, the small sample fallacy needs to be considered as a possible additional complication.

Conjunction fallacy

The conjunction fallacy is produced by pairing a less likely event with a more likely event. Tversky and Kahnman's (1982b, p. 90) heuristic bias is based on the belief that the conjunction is more representative and so more probable than is the less likely event by itself. Here representativeness can be said to be defined along the dimension of the more likely event. The normative rule is that the conjunction must be less probable than either event by itself, because the extra specification of the conjunction reduces the number of instances.

The causal conjunction fallacy is a special case of the conjunction fallacy. A possible event is judged to be more likely when it is combined with a plausible causal event. However, Tversky and Kahneman (1983, p. 308) suggest that the causal conjunction encourages people to judge the probability of the event given the cause, p(event/cause). They may not judge the probability of the conjunction, p(event and cause) as they should do.

Regression fallacy

The regression fallacy occurs with repeated measures. The normative rule is that when a recent average score happens to lie well above or well below the true average, future scores will regress towards the true average. Kahneman and Tversky (1973, p. 250) attribute the fallacy to the heuristic that future scores should be maximally representative of past scores, and so should not regress. Representativeness can be defined along any dimension that shows regression.

Base rate neglect

A base rate is the prior probability of a characteristic in a population. Base rate neglect (Kahneman and Tversky, 1973, p. 237) occurs when there are 2 independent probabilities of an event, a base rate or prior probability and a likelihood probability. The normative rule is to combine the 2 independent probabilities using the Bayes product of odds method:

$$\text{Prior odds} \times \text{likelihood ratio} = \text{posterior odds} \qquad (1.1)$$

The heuristic bias is to ignore or to give less weight to the prior probability or base rate because it is judged to be less representative or individuating than is the likelihood probability.

In the extreme form of base rate neglect, the prior probability or base rate is completely ignored. However, completely ignoring the base rate can be an incidental consequence of making an undetected logical fallacy (Braine, Connell, Freitag and O'Brien, 1990).

Availability and simulation fallacies

Tversky and Kahneman's (1973, pp. 208–9; Kahneman and Tversky, 1982c, Chapter 14) availability and simulation heuristics are used when people do not know the frequency or probability of instances in the outside world, and so cannot follow the normative rule of using objective measures. Instead they have to use, as a substitute, the heuristic of judging the frequency or probability from the availability of instances in their memory or from the ease with which they can simulate or imagine instances. When memory or ease of simulation does not provide a reasonable estimate of objective availability, they commit the availability or simulation fallacy.

Anchoring and adjustment biases

When people have an obvious anchor in dealing with probabilities, they can use Tversky and Kahneman's (1974, p. 1128) anchoring and adjustment heuristic, instead of following the normative rule of avoiding the use of anchors. In one version of the anchoring and adjustment heuristic,