# CHAPTER 10

# THE REPRESENTATIVENESS HEURISTIC

How do people come to their decisions? How do they choose among different options? And how do they form judgments of the value or likelihood of particular events or outcomes? This section of the book focuses on two related issues: the *processes* by which decision makers reach their conclusions, and the *biases* that can result as a consequence of these processes.

Amos Tversky and Daniel Kahneman (1974) have proposed that decision makers use "heuristics," or general rules of thumb, to arrive at their judgments. The advantage of heuristics is that they reduce the time and effort required to make reasonably good judgments and decisions. For example, it is easier to estimate how likely an outcome is by using a heuristic than by tallying every past occurrence of the outcome and dividing by the total number of times the outcome could have occurred. In most cases, rough approximations are sufficient (just as people often satisfice rather than optimize).

Normally, heuristics yield fairly good estimates. The disadvantage of using heuristics, however, is that there are certain instances in which they lead to systematic biases (i.e., deviations from normatively derived answers). The heuristic discussed in this chapter is known as "representativeness," and it leads to very predictable biases in certain situations. As mentioned earlier, the reason for focusing on biases rather than successes is that biases usually reveal more of the underlying processes than do successes. In fact, virtually all current theories of decision making are based on the results of research concerning biases in judgment.

## THE A, B, C's OF REPRESENTATIVENESS

According to Tversky and Kahneman (1974, p. 1124), people often judge probabilities "by the degree to which A is representative of B, that is, by the degree to which A resembles B." Tversky and Kahneman called this rule of thumb the "representativeness heuristic."

What are "A" and "B"? It depends on the judgment you are making. If you are estimating the probability that A came from B, then A might be an instance or a sample, and B might be a category or a parent population. For example, A might be a person, B might be a group, and the judgment in question might be the probability that A is a member of B. On the other hand, if you are trying to estimate the probability that A was produced by B, then A might be an event or an effect, and B might be a process or cause. For instance, B might be the process of flipping an unbiased coin, A might be the event of getting six Heads in a row, and the judgment might concern the chances of observing such an event with an unbiased coin. Because this definition of representativeness is abstract and a little hard to understand, let's consider some concrete examples of how the representativeness heuristic works and how it can lead to biases in certain situations.

Item #1 of the Reader Survey provides one example. This problem, taken from a study by Tversky and Kahneman (1982), reads as follows:

> 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 antinuclear demonstrations. Please check off the most likely alternative:
>
> ❑ Linda is a bank teller.
> ❑ Linda is a bank teller and is active in the feminist movement.

Most people feel that Linda is more likely to be a feminist bank teller than a bank teller. When Tversky and Kahneman (1982) put this question to 86 people, nearly 9 of every 10 respondents answered this way. If you think about it, though, this response violates a fundamental rule of probability. The conjunction, or co-occurrence, of two events (e.g., "bank teller" and "feminist") cannot be more likely than the probability of either event alone (e.g., "bank teller"). For this reason, Tversky and Kahneman (1983) called this phenomenon the "conjunction fallacy" (see also Leddo, Abelson, & Gross, 1984; Morier & Borgida, 1984).

You can verify the conjunction rule by looking at Figure 10.1. The circle on the left represents the universe of all bank tellers, the circle on the right represents the universe of all feminists, and the shaded area represents all bank tellers who are feminists. Because some bank tellers are *not* feminists, the chances of being a bank teller (whether feminist or not) will always be greater than the chances of being a bank teller who is feminist.

Just to make sure that people were not interpreting "bank teller" to mean "bank teller who is not active in the feminist movement," Tversky and Kahneman (1982) ran additional experiments in which different groups of subjects were presented with a set of alternatives that included one of the alternatives from Item #1 but not the other (so that the two alternatives were never directly compared). Even in these experi-

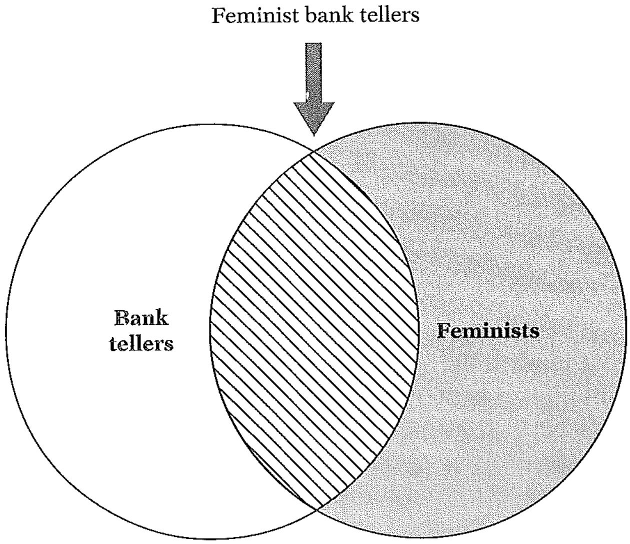


FIGURE 10.1
The overlapping worlds of bank tellers and feminists.

ments, subjects assigned a higher probability to Linda being a feminist bank teller than to Linda being a bank teller.

Tversky and Kahneman (1982) also found similar results with problems about "Bill" (who was thought more likely to be an accountant and jazz player than simply a jazz player), a Wimbledon tennis player (who was thought more likely to lose the first set but win the match than simply to lose the first set), and a former president of the United States (who was thought more likely to provide federal support for unwed mothers and to cut federal support to local governments than simply to provide federal support for unwed mothers).

From results such as these, Tversky and Kahneman (1982, p. 98) concluded: "As the amount of detail in a scenario increases, its probability can only decrease steadily, but its representativeness and hence its apparent likelihood may increase. The reliance on representativeness, we believe, is a primary reason for the unwarranted appeal of detailed scenarios and the illusory sense of insight that such constructions often provide. . . . For example, the hypothesis 'the defendant left the scene of the crime' may appear less plausible than the hypothesis 'the defendant left the scene of the crime for fear of being accused of murder,' although the latter account is less probable than the former."

This conclusion is further supported by the way that most people respond to Item #11 of the Reader Survey. This question asked which of the following two scenarios was more likely:

| | |
|---|---|
| **Scenario 1:** | An all-out nuclear war between the United States and Russia |
| **Scenario 2:** | A situation in which neither country intends to attack the other side with nuclear weapons, but an all-out nuclear war between the United States and Russia is triggered by the actions of a third country such as Iraq, Libya, Israel, or Pakistan |

As with the bank teller problem, most people feel that the more specific event (an all-out war triggered by a third country) is more probable than the more general event (an all-out war). Indeed, the Pentagon has spent decades developing war plans and procuring weapons to handle highly detailed but extremely improbable scenarios. According to Tversky and Kahneman, specific scenarios appear more likely than general ones because they are more representative of how we imagine particular events.

## THE LAW OF SMALL NUMBERS

Another consequence of the representativeness heuristic is that people tend to believe in what Tversky and Kahneman (1971) call "the law of small numbers." The law of small numbers is a tongue-in-cheek reference to a law in statistics known as the law of large numbers (a law stating that the larger a sample you draw from a population, the closer its average will be to the population average). A belief in the law of *small* numbers is a belief that random samples of a population will resemble each other and the population more closely than statistical sampling theory would predict.

For example, when people are asked to write down a random sequence of coin tosses without actually flipping a coin, they often try to make the string look random at every point (Kahneman and Tversky, 1972, call this "local representativeness"). As a consequence, they tend to exclude long runs and include more alternations between Heads and Tails than you would normally find in a chance sequence. In a chance sequence, there are many points at which the series does not look random at all. To verify this fact, you can approximate a chance sequence by tossing a coin 100 times and recording the pattern of Heads and Tails.

An illustration of the law of small numbers is given in Item #15 of the Reader Survey. This problem comes from a study by Tversky and Kahneman (1971), and it runs as follows:

> The mean IQ of the population of eighth graders in a city is *known* to be 100. You have selected a random sample of 50 children for a study of educational achievements. The first child tested has an IQ of 150. What do you expect the mean IQ to be for the whole sample?

Most people answer that the mean IQ should still be 100, but in fact, the correct answer is that the average IQ should be 101. The correct answer is 101 because the first child has an IQ of 150 and the 49 remaining children have an expected IQ of 100 each. This makes a total of 5050 IQ points (150 + 4900) which, when divided by 50 children, comes out to an average expected IQ of 101.

If you answered 100 rather than 101, you probably assumed that there would be low IQ scores to "balance out" the high score of 150. Such a view implicitly assumes, however, that chance is self-correcting. Chance does *not* correct or cancel out high scores with correspondingly low scores; it merely "dilutes" high scores with additional scores that are closer to the average (in this case 100). Tversky and Kahneman (1971) have argued that the tendency to view chance as self-correcting is an example of a bias resulting from the representativeness heuristic, because samples are expected to be highly representative of their parent population.

In the same vein, Tversky and Kahneman (1971) proposed that the representativeness heuristic leads people to commit the "gambler's fallacy"—the belief that a successful outcome is due after a run of bad luck (or, more generally, the belief that a series of independent trials with the same outcome will soon be followed by an opposite outcome). Item #31 of the Reader Survey examined your tendency to commit the gambler's fallacy. You were asked:

> Suppose that an unbiased coin is flipped three times, and each time the coin lands on Heads. If you had to bet $100 on the next toss, what side would you choose?

Because the coin is unbiased, the normatively correct answer is that you should have no preference between Heads and Tails. Some people erroneously believe, however, that Tails is more probable after a run of three Heads. Tversky and Kahneman explain this answer in terms of the mistaken belief that chance sequences must be locally representative (i.e., that every part of the sequence must appear random).

## THE HOT HAND

One of the most entertaining demonstrations of the law of small numbers was published by Thomas Gilovich, Robert Vallone, and Amos Tversky (1985). Instead of looking at coin tosses, these researchers examined people's perceptions of a "hot hand" in basketball. A player with a hot hand (also known as a "streak shooter" or an athlete "on a roll") is a player who has a better chance of making a basket after one or more successful shots than after having missed a shot.

What Gilovich, Vallone, and Tversky discovered is that Philadelphia 76er basketball fans—and several players and coaches as well—perceived streak shooting when statistical analyses showed that none existed. That is, people thought that the chances of making a basket increased after a player had made several successful shots, when in truth the chances of making the next basket were not significantly different from the player's overall probability of making a basket. Gilovich and his associates found the same results with free-throw records from the Boston Celtics and with laboratory experiments (or, more precisely, *gymnasium* experiments) on men and women from Cornell's varsity basketball teams.

For a short time, these findings created a national uproar in the sports community. How could Gilovich, Vallone, and Tversky say that streak shooting was simply an illusion? Anyone who has played or watched basketball *knows* that players are sometimes hot or cold! Basketball teams even alter their strategies in order to defend against streak shooters. The idea that basketball shots made by the same player are statistically unrelated seems very hard to swallow.

In order to find out why people strongly perceive streak shooting when successes and failures are statistically independent of one another, Gilovich, Vallone, and Tversky (1985) conducted an experiment in which subjects viewed six different series of X's and O's (which you might think of as "hits" or "misses" in a basketball game). Each series contained 11 X's and 10 O's, and the probability of alternating between the two letters was set at .40, .50, .60, .70, .80, or .90. For example, the string "XOXOXOOOXXOXOXOOXXXOX" represented an alternation probability of .70 (because it alternated between X and O on 14 of 20 possible alternations).

Gilovich, Vallone, and Tversky found that subjects selected the .70 and .80 sequences as the best examples of a chance series, rather than correctly selecting the .50 sequence. The sequence with a .50 probability of alternation was classified as a chance series by only 32 percent of the subjects. Indeed, 62 percent of the subjects classified the .50 sequence as "streak shooting."

To see how you would have performed on this task, take a look at your answer to Item #38 of the Reader Survey. The first string (XOXXXOOOOXOXXOOOXXXOX) alternates on half of all possible occasions (similar to what might be expected from a chance series). In contrast, the second sequence (XOXOXOOOXXOXOXOOXXXOX) represents an alternation probability of .70—far higher than the .50 expected by chance alone. If you thought that the first sequence contained runs that were too long to have been generated randomly, you were expecting too many alternations between X and O in the series (just as people do when they see "streak shooting" in chance sequences). Chapter 14 discusses the perception of randomness in greater detail.

## NEGLECTING BASE RATES

In some instances, a reliance on representativeness leads people to ignore "base rate" information (a base rate is the relative frequency with which an event occurs). Kahneman and Tversky have demonstrated this tendency in a series of experiments. In one study, for example, Kahneman and Tversky (1973, p. 241) told subjects that:

> A panel of psychologists have [sic] interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.

For example, here is a thumbnail description that Kahneman and Tversky intended to be fairly representative of an engineer:

> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

Using the same five thumbnail descriptions, Kahneman and Tversky also gave a second group of subjects identical instructions with the proportion of engineers and lawyers reversed (that is, 70 engineers and 30 lawyers), but because the results are comparable, we will focus exclusively on the condition with 30 engineers.

Once subjects rated the probability that each of the five people might be an engineer, they were asked to estimate the probability that someone randomly selected from the pool of 100 descriptions (a person about whom they were given no information) would be an engineer. Not surprisingly, on average subjects rated the chances that a randomly chosen person would be an engineer as roughly 30 percent. In other words, they used the base rate given in the problem.

On the other hand, when subjects were provided with descriptive information—even information that had nothing to do with being an engineer or lawyer—they tended to ignore base rates. For example, Kahneman and Tversky (1973) deliberately constructed the following portrait to be equally descriptive of an engineer or a lawyer:

> 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.

Such a description is entirely uninformative with respect to Dick's profession; consequently, the probability of being an engineer in this case should be equal to the base rate of 30 percent. Kahneman and Tversky (1973) found, however, that subjects given this description gave a median probability estimate of 50 percent. Apparently, subjects ignored the base rate information and simply judged the description as equally representative of an engineer or a lawyer.

A good deal of research has investigated the conditions under which people tend to use or neglect base rate information (Bar-Hillel, 1980, 1990; Fischhoff & Bar-Hillel, 1984; Osberg & Shrauger, 1986). For example, Icek Ajzen (1977) found that people often use base rate information when it is consistent with their intuitive theories of cause and effect. In one experiment, Ajzen asked subjects to predict a student's grade point average based on either causal factors (such as the number of hours per week the student studied) or noncausal information (such as the student's weekly income). Ajzen found that people used base rates more often when the information was causal than when it was not—even when they were told that the noncausal factors predicted grade point average just as well as the causal factors.

## NONREGRESSIVE PREDICTION

People also tend to neglect the diagnosticity of the information on which they base their predictions, and as a result, they make "nonregressive" predictions. For example, Item #35 of the Reader Survey (based on a problem from Kahneman and Tversky, 1973) posed the following question:

> Suppose that scores on a high school academic achievement test are moderately related to college grade point averages (GPAs). Given the percentiles below [shown in Figure 10.2], what GPA would you predict for a student who scored 725 on the test?

How did you respond? Most people predict a GPA between 3.5 and 3.7 (a GPA highly "representative" of a 725 test score). This answer makes sense if the achievement test is perfectly diagnostic of a student's GPA. If there is an exact correspondence between test scores and GPAs, then a score of 725 would translate into a GPA of about 3.6. According to the problem, though, scores on the achievement test are *not* perfect predictors of GPA. The problem states that scores on the test are only "moderately related to college grade point average." Because test scores are only moderately predictive of GPA, the best GPA prediction lies between 3.6 and the average GPA of 2.5—thereby allowing for "regression to the mean."

FIGURE 10.2
The relationship between performance on an achievement test and grade point average. *(Taken from Item #35 of the Reader Survey.)*

| Student Percentile | Achievement test | GPA |
|---|---|---|
| Top 10% | > 750 | > 3.7 |
| Top 20% | > 700 | > 3.5 |
| Top 30% | > 650 | > 3.2 |
| Top 40% | > 600 | > 2.9 |
| Top 50% | > 500 | > 2.5 |

Regression to the mean is a statistical phenomenon in which high or low scores tend to be followed by more average scores, just as very tall parents tend to have children who are closer to average height. Because a test score of 725 is quite high, a student who is tested a second time would probably receive a score that is somewhat closer to the average of 500 (and, by the same logic, would have a correspondingly lower GPA). You can think of it this way: A GPA of 2.5 is the best guess if you have no information about the student, and a GPA of 3.6 is the best guess if high school achievement test scores correlate perfectly with college GPAs. Because test scores are only *moderately* predictive of GPAs, the best choice is somewhere between 2.5 and 3.6 (i.e., somewhat higher than the average, but not nearly as high as 3.6).

Most psychologists think of test scores as being made up of two independent components: the "true score" and "error." The true score is the score that a student would receive if the test were a perfect measure of ability, and the error component is the result of all the factors that have nothing to do with ability but nonetheless influence a particular test score (amount of sleep the previous night, blood sugar level, mood, lighting conditions, and so on). In most cases these factors tend to cancel each other out, but occasionally they combine so as to dramatically increase or decrease a test score. Because this fluctuation is independent of the true score, however, future test scores are likely to "regress" toward the true score.

The tendency to overlook regression can lead to critical errors in judgment. For example, Kahneman and Tversky (1973) discuss a case in which instructors in a flight school concluded that praising pilots for well-executed flight maneuvers caused a decline in subsequent performance. Does this decline mean that teachers should stop reinforcing successes? Not at all! On the basis of regression alone, outstanding performances will be followed by performances that are closer to the average. Likewise, poor performances that are punished will improve regardless of whether punishment is genuinely effective.

In their book on human inference, Richard Nisbett and Lee Ross (1980, pp. 163, 165) describe some additional consequences of misinterpreting regression:

> Such mislabeling of simple regression phenomena (whereby extremely good or bad performances will, on the average, be followed by less extreme performances whenever there is an element of chance in such performances) is common in everyday experience. One disconcerting implication of such mislabeling is that measures designed to stem a "crisis" (a sudden increase in crime, disease, or bankruptcies, or a sudden decrease in sales, rainfall, or Olympic gold medal winners) will, on the average, seem to have greater impact than there actually has been. . . . Superstitions about what one must change to end a "bad streak" of outcomes, or must *not* change for fear of ending a "good streak," will arise from the observation of simple regression phenomena.

George Gmelch (1978, August), a professional baseball player who later became a social science researcher, chronicled several examples of such superstition in an article entitled "Baseball Magic." According to Gmelch, the New York Giants refused to clean their uniforms during a 16-game winning streak for fear of "washing away" their good fortune. Similarly, Leo Durocher wore the same black shoes, grey slacks, blue coat, and knotted tie for three and a half weeks while the Brooklyn Dodgers clinched a pennant victory in 1941.

Regression toward the mean can also explain why highly successful athletes and teams tend to experience a drop in performance immediately after appearing on the cover of *Sports Illustrated* magazine. Athletes typically appear on the cover following an unusually good performance, and from regression alone, a decline in performance would be expected. The "*Sports Illustrated* Jinx," as it is known, is not a jinx at all—most likely, it is nothing more than regression to the mean.

## CLINICAL VERSUS ACTUARIAL PREDICTION

The tendency people have to neglect information on base rates and diagnosticity contributes to a very surprising and embarrassing state of affairs. As documented in nearly 100 studies in the social sciences (Dawes, Faust, & Meehl, 1989), the accuracy of "actuarial" predictions (predictions based solely on empirical relations between a given set of variables and an outcome) is equal to or better than the accuracy of "clinical" predictions (predictions based on the judgment of human beings). In other words, contrary to common sense, predictions are usually more accurate when they are *not* made by a human decision maker—even when the decision maker has full access to actuarial information.

For example, in one study on clinical prediction, the judgments of 21 psychiatric staff members were compared with the weight of patients' compensation claim files in predicting readmission for psychiatric care (Lasky, Hover, Smith, Bostian, Duffendack, & Nord, 1959). The weight of claim files was used as a crude measure of past hospitalizations. As it turned out, staff judgments were not significantly more accurate in predicting readmission than were folder weights (the correlations were .62 and .61, respectively).

Apparently, the expertise of staff members and the advantage of having additional information were more than offset by other factors. Because clinical judges generally rely on heuristics such as representativeness—and are therefore susceptible to a variety of biases—their predictions are rarely more accurate than predictions based exclusively on actuarial relations.

## CONCLUSION

Research on the representativeness heuristic suggests several ways to improve judgment and decision making skills, including the following tips:

- ***Don't Be Misled by Highly Detailed Scenarios.*** The very specificity that makes detailed scenarios seem representative also lessens their likelihood. In general, the more specific a scenario is, the lower its chances are of occurring—even when the scenario seems perfectly representative of the most probable outcome.
- ***Whenever Possible, Pay Attention to Base Rates.*** Base rates are particularly important when an event is very rare or very common. For example, because the base rate is so low, many talented applicants never get admitted to graduate school (and it would be a mistake to interpret nonadmission as indicating that an applicant lacks academic ability). Conversely, because the base rate is so high, many unskilled drivers are awarded a driver's license. When base rates are extreme, representativeness is often a fallible indicator of probability.
- ***Remember That Chance Is Not Self-Correcting.*** A run of bad luck is just that: a run of bad luck. It does not mean that an equivalent run of good luck is bound to occur, and it does not mean that things are doomed to stay the same. If a chance process (like tossing an unbiased coin) has a certain probability of producing a certain outcome, past events will have no effect on future outcomes.
- ***Don't Misinterpret Regression Toward the Mean.*** Even though a run of bad luck is not necessarily balanced by a run of good luck (or vice versa), extreme performances tend to be followed by more

average performances. Regression toward the mean is typical whenever an outcome depends in part upon chance factors. Every once in awhile these factors combine to produce an unusual performance, but on subsequent occasions, performance usually returns to normal.

By keeping these suggestions in mind, it is possible to avoid many of the biases that result from a reliance on the representativeness heuristic. Chapter 11 focuses on another well-known heuristic, "availability," and the biases that often result from it.