CHAPTER 10

THE ACHILLES' HEEL OF HUMAN COGNITION

PROBABILISTIC REASONING

Question: Men are taller than women, right?

Answer: *"Right."*

Question: All men are taller than all women, right?

Answer: *"Wrong."*

Correct. Believe it or not, we are going to devote part of this chapter to something that you just demonstrated you knew by answering the previous two questions. But don't skip this chapter just yet, because there are some surprises waiting in the explanation of what seems like a very simple principle.

You answered affirmatively to the first question because you did not interpret "Men are taller than women" to mean what the second statement said: "All men are taller than all women." You correctly took the first statement to mean *"There is a tendency for* men to be taller than women," because everyone knows that not all men are taller than all women. You correctly interpreted the statement as reflecting a probabilistic trend rather than a fact that holds in every single instance. By *probabilistic trend,* we simply mean that it is more likely than not but does not hold true in all cases. That is, the relationship between sex and height is stated in terms of likelihoods and probabilities rather than certainties. Many other relationships in nature are probabilistic: It tends to be warmer near the equator. Families tend to have fewer than eight children. Most parts of the earth tend to have more insects than humans. These are all statistically demonstrable trends, yet there are exceptions to every one of them. They are probabilistic trends and laws, not relationships that hold true in every single case.

Virtually all the facts and relationships that have been uncovered by the science of psychology are stated in terms of probabilities. There is nothing unique about this. Many of the laws and relationships in other sciences are stated in probabilities rather than certainties. The entire subdiscipline of population genetics, for example, is based on probabilistic relationships. Physicists tell us that the distribution of the electron's charge in an atom is described by a probabilistic function.

It is true that most of the probabilistic trends uncovered in psychology are weaker than those in other sciences. However, the fact that behavioral relationships are stated in probabilistic form does not distinguish them from those in many other sciences. Jacob Bronowski (1978a) discussed the difficulties that many people have in accepting the idea that, as science progresses into new areas, more and more of its laws are stated in probabilistic terms:

> If I say that after a fine week, it always rains on Sunday, then this is recognized and respected as a law. But if I say that after a fine week, it rains on Sunday more often than not, then this somehow is felt to be an unsatisfactory statement; and it is taken for granted that I have not really got down to some underlying law which would chime with our habit of wanting science to say decisively either "always" or "never." Even if I say that after a fine week, it rains on seven Sundays out of ten, you may accept this as a statistic, but it does not satisfy you as a law. Somehow it seems to lack the force of a law. Yet this is a mere prejudice.... The idea of chance as I have explained it here is not difficult. But it is new and unfamiliar. We are not used to handling it.... We seem to be in a land of sometimes and perhaps, and we had hoped to go on living with always and with certainty.... Yet I believe that the difficulty is only one of habit. We shall become accustomed to the new ideas just as soon as we are willing and as we have to. And we are having to. (pp. 81–82, 94–95)

In this chapter, we will try to make you more comfortable in the "land of sometimes and perhaps," because to understand psychology one must be comfortable with the subject of this chapter: probabilistic reasoning.

"PERSON-WHO" STATISTICS

Most of the public is aware that many of the conclusions of medical science are statements of probabilistic trends and are not predictions of absolute certainty. Smoking causes lung cancer and a host of other health problems. Voluminous medical evidence documents this fact (Brandt, 1990; Jeffrey, 1989). Yet will everyone who smokes get lung cancer, and will everyone who refrains from smoking be free of lung cancer? Most people know that these implications do not follow. The relationship is probabilistic. Smoking vastly increases the *probability* of contracting lung cancer but does not make it a certainty. Medical science can tell us with great confidence that more people in a group of smokers will die of lung cancer than in an equivalent group of nonsmokers. It cannot tell us *which* ones will die, though. The relationship is probabilistic; it does not hold in every case. We are all aware of this—or are we? How often have we seen a nonsmoker trying to convince a smoker to stop by citing the smoking—lung-cancer statistics, only to have the smoker come back with "Oh, get outta here! Look at old Joe Ferguson down at the store. Three packs of Camels a day since he was sixteen! Eighty-one years old and he looks great!" The obvious inference that one is supposed to draw is that this single case somehow invalidates the relationship.

It is surprising and distressing how often this ploy works. Too frequently, a crowd of people will begin to nod their heads in assent when a single case is cited to invalidate a probabilistic trend. This agreement reflects a failure to understand the nature of statistical laws. If people think a single example can invalidate a law, they must feel the law should hold in every case. In short, they have failed to understand the law's probabilistic nature. There will always be a few "people who" go against even the strongest of trends. Consider our smoking example. Only 5 percent of men who live to the age of 85 are smokers (University of California, Berkeley, 1991). Or to put it another way, 95 percent of men who live to age 85 are either nonsmokers or have smoked for a period and then quit. Continuous smoking without quitting markedly shortens lives (University of California, Berkeley, 1991). Yet a few smokers do make it to 85.

Adapting the terminology of psychologists Richard Nisbett and Lee Ross (1980), we will call instances like the "Joe Ferguson" story examples of the use of "person-who" statistics: situations in which well-established statistical trends are questioned because someone knows a "person who" went against the trend. For example, "You say job opportunities are expanding in service industries and contracting in heavy industry? No way. I know a man who got a job in a steel mill just last Thursday"; "You say families are having fewer children than they did 30 years ago? You're crazy! The young couple next door already has 3 and they're both under 30"; "You say children tend to adopt the religious beliefs of their parents? Well, I know a man at work whose son converted to another religion just the other day"; "You say the average food-stamp recipient receives a subsidy averaging 49 cents per meal? Yeah, but what about a woman I saw at a store who purchased lots of expensive meat with her stamps."

The ubiquitous person-who is usually trotted out when we are confronted with hard statistical evidence that contradicts a previously held belief. Thus, it could be argued that people actually know better and simply use the person-who as a technique to invalidate facts that go against their opinions. However, the work of psychologists who have studied human decision making and reasoning suggests that the tendency to use the person-who comes not simply from its usefulness as a debating strategy. Instead, it appears that this fallacious argument is used so frequently because people experience great difficulty in dealing with probabilistic information. New work in the psychology of decision making has indicated that probabilistic reasoning may well be the Achilles' heel of human cognition.

PROBABILISTIC REASONING AND THE MISUNDERSTANDING OF PSYCHOLOGY

Probabilistic thinking is involved in many areas of science, technology, and human affairs. Thus, there is no necessary reason why this type of thinking is more important to an understanding of psychology. However, the findings of psychology are quite often misunderstood because of the problems people have in dealing with probabilistic information. We all understand "Men are taller than women" as a statement of probabilistic tendency. We realize that it is not invalidated by a single exception (one man who is shorter than a woman). Most people understand the statement "Smoking causes lung cancer" in the same way, although old "Joe Ferguson" can be convincing to some smokers who do not want to believe that their habit may be killing them. However, very similar probabilistic statements about *behavioral* trends cause widespread disbelief and are often dismissed by many people with the first appearance of a single person-who. Most psychology instructors have witnessed a very common reaction when they discuss the evidence on certain behavioral relationships. For example, the instructor may present the fact that children's scholastic achievement is related to the socioeconomic status of their households and to the educational level of their parents. This statement often prompts at least one student to object that he has a "friend" who is a National Merit Scholar and whose father finished only eighth grade. Even those who understood the smoking–lung-cancer example tend to waver at this point.

People who would never think of using person-who arguments to refute the findings of medicine and physics routinely use them to refute psychological research. Most people understand that many treatments, theories, and facts developed by medical science are probabilistic. They understand that, for example, a majority of patients, but not all of them, will respond to a certain drug. Medical science, however, often cannot tell in advance *which* patients will respond. Often all that can be said is that if 100 patients take treatment A and 100 patients do not, after a certain period the 100 patients who took treatment A will *collectively* be better off. No one would doubt the worth of this medical knowledge just because it is probabilistic and does not apply in every case. Yet this is exactly what happens in the case of many psychological findings and treatments. The fact that a finding or treatment does not apply in every case often engenders profound disappointment and denigration of psychology's progress. When the issues are psychological, people tend to forget the fundamental principle that knowledge does not have to be certain to be useful—that even though individual cases cannot be predicted, the ability to accurately forecast group trends is often very informative. The prediction of outcomes based on group characteristics is often called *aggregate* or *actuarial prediction* (we will discuss actuarial prediction in more detail in the next chapter).

For these reasons, a thorough understanding of probabilistic reasoning is critical to an understanding of psychology. There is a profound irony here. Psychology probably suffers the most from the general public's inability to think statistically. Yet psychologists have done the most research into the nature of probabilistic reasoning abilities.

PSYCHOLOGICAL RESEARCH ON PROBABILISTIC REASONING

In the past two decades, the research of psychologists such as Daniel Kahneman of Princeton University, Richard Nisbett at the University of Michigan, and the late Amos Tversky has revolutionized the way we think about people's reasoning abilities. In the course of their studies, these investigators have uncovered some fundamental principles of probabilistic reasoning that are absent or, more commonly, insufficiently developed in many people. As has often been pointed out, it should not be surprising that they are insufficiently developed. As a branch of mathematics, statistics is a very recent development (Hacking, 1975). Games of chance existed centuries before the fundamental laws of probability were discovered. Here is another example of how personal experience does not seem to be sufficient to lead to a fundamental understanding of the world (see Chapter 7). It took formal study of the laws of probability to reveal how games of chance work: thousands of gamblers and their "personal experiences" were insufficient to uncover the underlying nature of games of chance.

The problem is that as society becomes more complex, the need for probabilistic thinking becomes greater for everyone. If ordinary citizens are to have a basic understanding of the society in which they live, they must possess at least a rudimentary ability to think statistically.

"Why did they raise my insurance rate," you might wonder, "and why is John's rate higher than Bill's? Is Social Security going broke? Is our state lottery crooked? Is crime increasing or decreasing? Why do doctors order all those tests? Why can people be treated with certain rare drugs in Europe and not in the United States? Do women really make less than men in comparable jobs? Do international trade deals cost Americans jobs and drive down wages? Is educational achievement in Japan really higher than here? Is Canadian health care better than that in the United States and cheaper as well?" These are all good questions—concrete, practical questions about our society and how it works. To understand the answers to each of them, one must think statistically.

Clearly, a complete discussion of statistical thinking is beyond the scope of this book. We will, however, briefly discuss some of the more common pitfalls of probabilistic reasoning. A good way to start developing the skill of probabilistic thinking is to become aware of the most common fallacies that arise when people reason statistically. In addition, many are particularly relevant to understanding the importance of psychological findings and theories.

Insufficient Use of Probabilistic Information

Consider the following problem, developed by Tversky and Kahneman (1982): A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city in which the accident occurred. You are given the following data:

1. 85 percent of the cabs in the city are Green and 15 percent are Blue.
2. A witness identified the cab as blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness had correctly identified each of the two colors 80 percent of the time. That is, the witness called about 80 percent of the Blue cabs blue, but called 20 percent of the Blue cabs green. The witness also called about 80 percent of the Green cabs green, but called 20 percent of the Green cabs blue.
3. What is the probability that the cab involved in the accident was Blue rather than Green?

The answer to this question is not merely a matter of opinion. There is a specific rule of probability, Bayes's theorem, that dictates how the probability assessment in such a question is to be calculated. The theorem provides the optimal way of combining the two pieces of information that have been given: overall, 15 percent of the cabs are Blue and a witness, whose identification accuracy is 80 percent, identified the cab in question as Blue. Most people do not naturally combine the two pieces of information optimally. In fact, many people are surprised to learn that the probability that the cab is Blue is .41, and that, despite the witness's identification, it is still more likely that the cab involved in the accident was Green rather than Blue. The reason is that the general or prior probability that the cab is Green (85 percent) is higher than the credibility of the witness's identification of Blue (80 percent). Without using the formula, we can see how the probability of .41 is arrived at. In 100 accidents of this type, 15 of the cabs will be Blue and the witness would identify 80 percent of them (12) as Blue. Furthermore, out of 100 accidents of this type, 85 of the cabs will be Green and the witness will identify 20 percent of those 85 cabs (17) as Blue. Thus, 29 cabs will be identified as Blue, but only 12 of them will actually *be* Blue. The proportion of cabs identified as Blue that actually are Blue is 12 out of 29, or 41 percent.

Many people have been given the cab problem, and most estimates cluster in the range of .60–.80. That is, most people greatly overestimate the probability that the cab is Blue. They overweight the witness's identification and underweight the base rate, or prior probability, that the cab is Blue. This is another example of the tendency for concrete single-case information to overwhelm more abstract probabilistic information (the vividness problem discussed in Chapter 4). Of course, in this case, people seem to forget or ignore the fact that the identification by the witness is just as probabilistic as the base rate information.

One thing that is surprising is that the tendency to give insufficient weight to probabilistic information is not limited to the scientifically unsophisticated layperson. Casscells, Schoenberger, and Graboys (1978) gave a variant of the following problem to 20 medical students, 20 attending physicians, and 20 house officers at four Harvard Medical School teaching hospitals: Imagine that the virus (HIV) that causes AIDS occurs in 1 in every 1,000 people. Imagine also that there is a test to diagnose the disease that always indicates correctly that a person who has HIV actually has it. Finally, imagine that the test has a false-positive rate of 5 percent. This means that the test wrongly indicates that HIV is present in 5 percent of the cases in which the person does *not* have the virus. Imagine that we choose a person randomly and administer the test and that it yields a positive result (indicates that the person is HIV-positive). What is the probability that the individual actually has the HIV virus, assuming that we know nothing else about the individual's personal or medical history?

The most common answer was 95 percent. The correct answer is approximately 2 percent. The physicians vastly overestimated the probability that a positive result truly indicated the disease because of the same tendency to overweight the case information and underweight the base rate information that we discussed previously. Although the correct answer to this problem can be calculated by means of Bayes's rule, a little logical reasoning can help to illustrate the profound effect that base rates have on probabilities. Of 1,000 people, 1 will actually be HIV-positive. If the other 999 (who do not have the disease) are tested, the test will indicate incorrectly that approximately 50 of them have the virus (.05 multiplied by 999) because of the 5 percent false-positive rate. Thus, of the 51 patients testing positive, only 1 (approximately 2 percent) will actually be HIV-positive. In short, the base rate is such that the vast majority of people do not have the virus. This fact, combined with a substantial false-positive rate, ensures that, in absolute numbers, the vast majority of positive tests will be of people who do not have the virus.

Although all the physicians in the Casscells et al. study would have immediately recognized the correctness of this logic, their initial tendency was to discount the base rates and overweight the clinical evidence. In short, the physicians actually knew better but were initially drawn to an incorrect conclusion. Psychologists have termed problems like these *cognitive illusions*. In cognitive illusions, even when people know the correct answer, they may be drawn to an incorrect conclusion by the structure of the problem.

All the examples given here are cognitive illusions because they capitalize on a fallacy of human reasoning: the tendency to overweight individual-case evidence and underweight statistical information. The case evidence (the witness's identification, the laboratory test result) seems tangible and concrete to most people, whereas the probabilistic evidence seems, well—probabilistic. This reasoning, of course, is fallacious because case evidence itself is always probabilistic. A witness can make correct identifications with only a certain *degree* of accuracy, and a clinical test misidentifies the presence of a disease with a certain *probability*. The situation is one in which two probabilities—the probable diagnosticity of the case evidence and the prior probability—must be combined if one is to arrive at a correct decision. There are right and wrong ways of combining these probabilities, and more often than not—particularly when the case evidence gives the illusion of concreteness (recall our discussion of the vividness problem in Chapter 4)—people combine the information in the wrong way. This particular failure of probabilistic reasoning may very well hinder the use of psychological knowledge, which is often stated in the form of probabilistic relationships among behaviors.

Science writer K. C. Cole (1998) asks us to imagine two situations. One is the standard one in which we try to convey the dangers of smoking to an individual by stating a probability of death. The second way is more vivid. It asks the smoker to imagine that one pack in every 18,250 is different—it is filled with explosives. When the smoker opens it, the smokers dies. There is little doubt which imagine is most effective—yet they both are conveying *the same fact.*

Finally, the AIDS example is a good illustration of the importance of probabilistic thinking in our society. This was not a made-up problem. Mandatory AIDS testing in various employment and government settings has been a hotly debated issue for many years now, and sadly it will probably continue to be an issue in our society. As mathematics professor Lynn Steen (1990) argued:

> The continuing debate over mandatory AIDS testing provides a good example of quantitative issues hidden just beneath the surface of many public debates.... There will always be a small number (perhaps 2 percent) of errors.... The public infers from this that testing is 98 percent accurate. But since the actual incidence of AIDS in the general population is less than the error in the test, any widespread test administered to a random sample of citizens will produce more results indicative of AIDS because of errors in the test than the actual AIDS in the population. The personal consequences of these erroneous messages in psychological, economic, and emotional grief are rarely recognized by a public which naively assumes that any accurate test will produce accurate results when put into widespread use. (p. 218)

Inverting Conditional Probabilities

All dogs are mammals. Correct. All mammals are dogs. Incorrect. We realize the simple rule that if all A are B it does not necessarily follow that all B are A. Even with so simple a rule, though, things can sometimes get confusing. Consider the following two premises: All flowers have petals; roses have petals. Does it follow logically from these two premises that roses are flowers? At first glance, you might think yes. But look more closely. The statement "Roses are flowers" does *not* follow logically from the previous two premises, although even many college-educated subjects think it does (Markovits & Nantel, 1989; Sá, West & Stanovich, 1999; Stanovich, 1999). The error that even many educated subjects fall into is that they invert the first premise. They read all flowers have petals as implying that all things with petals are flowers. This inversion is what invites the inference that roses are flowers when one is given the premise that roses have petals. In fact, however, because some things with petals may not be flowers, the statement that roses have petals does not guarantee that roses are flowers.

Inverting premises is a common error of logical thinking. The mistaken inversion of statements happens even more frequently in the domain of probabilistic reasoning, particularly in reasoning about conditional probabilities. By *conditional probability,* we mean the probability of a particular event, A, given that another event, B, has happened. The inversion error in probabilistic reasoning is thinking that the probability of A, given B, is the same as the probability of B, given A. The two are not the same, yet they are frequently treated as if they are. Sometimes the difference between these two probabilities is easy to see because of the content of the problem. For example, it is obvious that the probability of pregnancy, given that intercourse has occurred, is different from the probability of intercourse, given that pregnancy has occurred!

However, sometimes the content of the problem leads us astray. For example, Dawes (1988) described an article in a California newspaper that ran a headline implying that most student users of marijuana were also using hard drugs. However, the first couple of lines of the article reported a survey indicating that most students who were using hard drugs had once smoked marijuana. The newspaper article had clearly inverted the probabilities. To put it in probabilistic terms, the headline implied that the survey was about the probability of using hard drugs, given the previous smoking of marijuana, but actually the article was about the inverse probability: the probability of having smoked marijuana, given that the student was using hard drugs. The problem is that the two probabilities are vastly different. The probability that students use hard drugs, given that they have smoked marijuana, is much, much smaller than the probability of having smoked marijuana given that students are using hard drugs. The reason is that most people who smoke marijuana do not use hard drugs, but most people who use hard drugs have tried marijuana.

Dawes (1988) cited the humorous example of a magazine article titled "This Quiz Could Save Your Life Next Weekend." The author of the article made the somewhat startling statement that "the farther one drives from home the safer one might be" (Dawes, 1988, p. 72). The basis of this statement was the statistic that most automobile deaths occur within 25 miles of home. In implying that if one drives farther from home, one will decrease the probability of death, the author had inverted probabilities. Just because there is a high probability that you will be within 25 miles of home if you are in a fatal car accident does *not* mean there is a high probability you will be in such an accident if you are within 25 miles of home.

Not so humorous, however, is a domain in which the inversion of conditional probabilities happens quite often: medical diagnosis (Dowie & Elstein, 1988; Eddy, 1982). It has been found that physicians (and sometimes their textbooks) tend to invert probabilities, thinking, mistakenly, that the probability of disease, given a particular symptom, is the same as the probability of the symptom, given the disease.

What if I told you that you had been given a cancer test and that the results were positive? Furthermore, what if I told you that this particular test had a diagnostic accuracy of 90 percent, that is, that 90 percent of the time, when cancer was present, this test gave a positive result? You might well be extremely upset. However, you might be much less upset if I told you that the chances that you had cancer were less than 50 percent. How can this be if the test has a diagnostic accuracy of 90 percent? Imagine that a study of this test was done in which 1,100 patients were tested, and that 100 of them actually had cancer. Imagine that the results were as follows:

	CANCER PRESENT	CANCER ABSENT
Test Positive	90	100
Test Negative	10	900

In this table, we can see the test's diagnostic accuracy of 90 percent. Of the 100 people with cancer, the test gave a positive result for 90. But you can immediately see that this is not the probability that is relevant to you. The 90 percent figure is the probability that the test result will be positive for someone who *has* cancer. But what you are interested in is the inverse: the probability that you have cancer, *given that* the test result is positive. And here the false alarm rate (10 percent—because the accuracy of the test is 90 percent) is relevant to you. Because of the false alarm rate of the test, 100 of the 1,000 people *without* cancer get a positive test result. A total of 190 people have gotten positive test results, but only 90 of them have cancer. Thus, the probability that you have cancer, *given that* the test result is positive is only 47.4 percent (90 divided by 190).

Unfortunately, this example is not just imaginary. Dawes (1988, p. 73) discussed a physician who was recommending a type of preventive treatment because he had confused the probability of cancer, given the diagnostic indicator, with the probability of the diagnostic indicator, given that the patient had cancer. Because the preventive treatment was a type of mastectomy, we can readily understand how serious this error in probabilistic reasoning can be.

Failure to Use Sample Size Information

Consider these two problems, developed by Tversky and Kahneman (1974):

1. A certain town is served by two hospitals. In the larger hospital, about 45 babies are born each day, and in the smaller hospital, about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it is higher than 50 percent, sometimes lower. For a period of one 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?
   a. The larger hospital
   b. The smaller hospital
   c. About the same
2. Imagine an urn filled with balls, two-thirds of which are of one color and one-third of which are of another. One individual has drawn 5 balls from the urn and found that 4 are red and 1 is white. Another individual has drawn 20 balls and found that 12 are red and 8 are white. Which of the two individuals should feel more confident that the urn contains two-thirds red balls and one-third white balls, rather than vice versa? What odds should each individual give?

In problem 1, the majority of people answer "about the same." People not choosing this alternative pick the larger and the smaller hospital with about equal frequency. Because the correct answer is the smaller hospital, approximately 75 percent of subjects given this problem answer incorrectly. These incorrect answers result from an inability to recognize the importance of sample size in the problem. Other things being equal, a larger sample size always more accurately estimates a population value. Thus, on any given day, the larger hospital, with its larger sample size, will tend to have a proportion of births closer to 50 percent. Conversely, a small sample size is always more likely to deviate from the population value. Thus, the smaller hospital will have more days on which the proportion of births displays a large discrepancy from the population value (60 percent boys, 40 percent boys, 80 percent boys, etc.).

In problem 2, most people feel that the sample of 5 balls provides more convincing evidence that the urn is predominantly red. Actually, the probabilities are in the opposite direction. The odds are 8 to 1 that the urn is predominantly red for the 5-ball sample, but they are 16 to 1 that the urn is predominantly red for the 20-ball sample. Even though the proportion of red balls is higher in the 5-ball sample (80 percent versus 60 percent), this is more than compensated for by the fact that the other sample is four times as large and thus is more likely to be an accurate estimate of the proportions in the urn. The judgment of most subjects, however, is dominated by the higher proportion of red in the 5-ball sample and does not take adequate account of the greater reliability of the 20-ball sample.

An appreciation of the influence of sample size on the reliability of information is a basic principle of evidence evaluation that applies in many different areas but, again, is particularly relevant to understanding research in the behavioral sciences. Whether we realize it or not, we all hold generalized beliefs about large populations. What we often do not realize is how tenuous the database is on which our most steadfast beliefs rest. Throw together some observations about a few neighbors and a few people at work, add some random anecdotes from the TV news, and we are all too ready to make statements about human nature or "the American people." This is to say nothing of the lack of representativeness of the sample, which is another issue entirely.

The Gambler's Fallacy

The gambler's fallacy is the tendency for people to see links between events in the past and events in the future when the two are really independent. Two outcomes are independent when the occurrence of one does not affect the probability of the other. Most games of chance that use proper equipment have this property. For example, the number that comes up on a roulette wheel is independent of the outcome that preceded it. Half the numbers on a roulette wheel are red, and half are black (for purposes of simplification, we will ignore the green zero and double zero), so the odds are even (.50) that any given spin will come up red. Yet after five or six consecutive reds, many bettors switch to black, thinking that it is now more likely to come up. This is the gambler's fallacy: acting as if previous outcomes affect the probability of the next outcome when the events are independent. In this case, the bettors are wrong in their belief. The roulette wheel has no memory of what has happened previously. Even if 15 reds in a row come up, the probability of red's coming up on the next spin is still .50.

You can demonstrate a similar phenomenon in beliefs among the general public about coin flipping. Ask a group of people what the probability of obtaining heads is on the sixth flip after five heads in a row have occurred, and some people will answer that it is very unlikely. This again is the gambler's fallacy. Coin flips are independent. After five consecutive heads, coins still have only two sides, each equally likely of turning up in any given trial.

The gambler's fallacy is not restricted to the inexperienced or novice gambler. Research has shown that even habitual gamblers, who play games of chance over 20 hours a week, still display belief in the gambler's fallacy (Wagenaar, 1988). Also, it is important to realize that the gambler's fallacy is not restricted to games of chance. It operates in any domain in which chance plays a substantial role, that is, in almost *everything*. The genetic makeup of babies is an example. Psychologists, physicians, and marriage counselors often see couples who, after having two female children, are planning a third child because "We want a boy, and it's *bound* to be a boy this time." This, of course, is the gambler's fallacy. The probability of having a boy (approximately 50 percent) is exactly the same after having two girls as it was in the beginning. The two previous girls make it *no more likely* that the third baby will be a boy.

The gambler's fallacy operates in any domain that has a chance component, such as sporting events and stock markets (see Andreassen, 1987). For example, one group of psychologists (Gilovich, Vallone, & Tversky, 1985) has studied the belief in "streak shooting" or the "hot hand" in basketball, that is, the belief that a particular shooter can "get hot" and that after making a series of shots, a player has a greater chance of making his next shot ("Get him the ball; he's hot"). The researchers ascertained that the belief in streak shooting was strong among both basketball fans and players. For example, on a questionnaire, 91 percent of a group of basketball fans believed that a player has a better chance of making a shot after having just made his last two or three shots than he does after having just missed his last two or three shots, and 84 percent of the fans believed that it is important to pass the ball to someone who has just made two or three shots in a row. The fans were asked to estimate, for a hypothetical player who shoots 50 percent from the field, what his field goal percentage would be after making a shot and what it would be after missing a shot. The fans estimated that after making a shot his percentage would be 61 percent and after missing a shot it would be 42 percent. The fans' strong beliefs in streak shooting were shared by most (but not all) of the players on the Philadelphia 76ers, who were interviewed by the researchers (see Gilovich et al., 1985).

But why are we discussing streak shooting here, under the heading of the gambler's fallacy? Because *there is no such thing as streak shooting!* Gilovich et al. (1985) studied the shooting statistics during the 1980–1981 season for the Philadelphia 76ers and the Boston Celtics. There were no sequential dependencies among the shots that the players took during the season. Let's see, nontechnically, what that means.

The gambler's fallacy is the belief that independent events are linked, that there are dependencies between events that are really not related. Statistically, the idea of streak shooting translates into the hypothesis that the probability of a hit (making a shot) after two or three consecutive hits is higher than the probability of a hit after two or three misses. Gilovich et al. (1985) calculated the probabilities and found that there was no evidence to support this hypothesis. For example, the data for Julius Erving (who took the most shots of anyone on the Philadelphia 76ers) showed that his probability of a hit after three consecutive hits was .48 and his probability of a hit after three previous misses was .52. Erving's probability of a hit after two consecutive hits was .52, compared to a probability of .51 after two consecutive misses. His probability of a hit after one previous hit was .53, compared to a probability of .51 after one miss. In short, Erving's probability of a hit was approximately .50 regardless of what had happened on his previous shots—no tendency toward streak shooting at all.

Data for other players were highly similar. Lionel Hollins had a field goal probability of .46 after two consecutive hits and a field goal probability of .49 after two consecutive misses. His probability of a hit after one hit was .46, exactly the same as his probability of a hit after one miss. Again, Hollins made approximately 47 percent of his shots regardless of what had happened on his previous shots. Data on free throws by the Boston Celtics showed the same thing. For example, Larry Bird made 88 percent of his free throws after making a free throw and made 91 percent of his free throws after missing a free throw. Nate Archibald made 83 percent of his free throws after making a free throw and made 82 percent of his free throws after missing a free throw. There are no streaks in free-throw shooting either. Belief in the "hot hand" is indeed an example of the gambler's fallacy, that is, believing that there are links among events that are really independent.

Interestingly, the gambler's fallacy appears to be another instance—like "intuitive physics," discussed in Chapter 6—in which mere experience does not reveal to people the true nature of the world. Gilovich et al. (1985) examined the performance of college basketball players who were practicing shots from about 15 feet on an open court (i.e., there were no defenders). They had the players make bets on 100 shots. The players were guaranteed to win because they had made about 50 percent of their shots from that distance, and the bets were structured so that the players won more when they made the shot than they lost when they missed the shot. However, the players could choose to make high bets (win a lot, lose a lot) or low bets (win a little, lose a little) before each shot. Obviously the players would do better if they could predict their own performance. That is, they should bet high when they thought the probability of making the shot was high and bet low when they thought the probability of making the shot was low. The results of this experiment indicated that, as with the professional players, there was no hot hand: The probability of making a shot after one or more made shots was no higher than the probability of making a shot after a miss. However, the players *thought* there was such a thing as a hot hand. They bet more on the next shot after they had made a shot than they bet after they had missed a shot. As a result, the players were *absolutely unable to predict their own performance:* their predictions were no better than chance.

The gambler's fallacy stems from many mistaken beliefs about probability. One is the belief that if a process is truly random, no sequence, not even a small one (six coin flips, for instance), should display runs or patterns. People routinely underestimate the likelihood of runs (HHHH) and patterns (HHTTHHTTHHTT) in a random sequence. For this reason, people cannot generate truly random sequences when they try to do so. The sequences that they generate tend to have too few runs and patterns. When generating such sequences, people alternate their choices too much in a mistaken effort to destroy any structure that might appear (Lopes & Oden, 1987).

Those who claim to have psychic powers can easily exploit this tendency. Consider a demonstration sometimes conducted in college psychology classes. A student is told to prepare a list of 200 numbers by randomly choosing from the numbers 1, 2, and 3 over and over again. After it is completed, the list of numbers is kept out of view of the instructor. The student is now told to concentrate on the first number on the list, and the instructor tries to guess what the number is. After the instructor guesses, the student tells the class and the instructor the correct choice. A record is kept of whether the instructor's guess matched, and the process continues until the complete record of 200 matches and nonmatches is recorded. Before the procedure begins, the instructor announces that she or he will demonstrate "psychic powers" by reading the subject's mind during the experiment. The class is asked what level of performance—that is, percentage of "hits"—would constitute empirically solid evidence of psychic powers. Usually a student who has taken a statistics course volunteers that, because a result of 33 percent hits could be expected purely on the basis of chance, the instructor would have to achieve a larger proportion than this, probably at least 40 percent, before one should believe that she or he has psychic powers. The class usually understands and agrees with this argument. The demonstration is then conducted, and a result of more than 40 percent hits is obtained, to the surprise of many.

The students then learn some lessons about randomness and about how easy it is to fake psychic powers. The instructor in this example merely takes advantage of the fact that people do not generate enough runs: They alternate too much when producing "random" numbers. In a truly random sequence of numbers, what should the probability of a 2 be after three consecutive 2s? One-third, the same as the probability of a 1 or a 3. But this is not how most people generate such numbers. After even a small run, they tend to alternate numbers in order to produce a representative sequence. Thus, on each trial in our example, the instructor merely picks one of the two numbers that the student did not pick on the previous trial. Thus, if on the previous trial the student generated a 2, the instructor picks a 1 or a 3 for the next trial. If on the previous trial the subject generated a 3, the instructor picks a 1 or a 2 on the next trial. This simple procedure usually ensures a percentage of hits greater than 33 percent—greater than chance accuracy—without a hint of psychic power.

A Further Word About Statistics and Probability

These, then, are just a few of the shortcomings in statistical reasoning that obscure an understanding of psychology. More complete and detailed coverage is provided in Nisbett and Ross's *Human Inference: Strategies and Shortcomings* *of Social Judgment* (1980) and in the book *Judgment Under Uncertainty: Heuristics and Biases* (1982), edited by Kahneman, Slovic, and Tversky. Popular introductions to many of these ideas (and good places to start for those who lack any mathematics training) are contained in Paulos's *Innumeracy: Mathematical Illiteracy and Its Consequences* (1988), in Gilovich's *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life* (1991), in Piattelli-Palmarini's *Inevitable Illusions: How Mistakes of Reason Rule Our Minds* (1994), and in Baron's (1998) *Judgment Misguided: Intuition and Error in Public Decision Making*. Somewhat more technical, but still very readable, introductions are also contained in Dawes's *Rational Choice in an Uncertain World* (1988) and in Baron's *Thinking and Deciding* (1994).

The probabilistic thinking skills discussed in this chapter are of tremendous practical significance. Because of inadequately developed probabilistic thinking abilities, physicians choose less effective medical treatments (McNeil, Pauker, Sox, & Tversky, 1982; Sutherland, 1992); people fail to accurately assess the risks in their environment (Margolis, 1996; Yates, 1992); information is misused in legal proceedings (Foster & Huber, 1999; Lees-Haley, 1997); millions of dollars are spent on unneeded projects by government and private industry (Dawes, 1988, pp. 23–24); animals are hunted to extinction (Gilovich, 1991, p. 5); unnecessary surgery is performed (Dawes, 1988, pp. 73–75); and costly financial misjudgments are made (Belsky & Gilovich, 1999; Fridson, 1993; Thaler, 1992; Willis, 1990).

Of course, a comprehensive discussion of statistical reasoning cannot be carried out in a single chapter. Our goal was much more modest: to emphasize the importance of statistics in the study and understanding of psychology. Unfortunately there is no simple rule to follow when confronted with statistical information. Unlike some of the other components of scientific thinking that are more easily acquired, functional reasoning skills in statistics probably require some type of formal study. Fortunately most universities and community colleges now offer introductory-level statistics courses that require no previous university-level mathematics. Before taking such a course, the reader can begin with the books I have recommended.

While many scientists sincerely wish to make scientific knowledge accessible to the general public, it is intellectually irresponsible to suggest that a deep understanding of a particular subject can be obtained by the layperson when that understanding is crucially dependent on certain technical information that is available only through formal study. Such is the case with statistics and psychology. Psychologist Alan Boneau (1990) surveyed authors of psychology textbooks, asking them to list the most important terms and concepts that students need to learn in psychology. Approximately 40 percent of the 100 terms and concepts that were listed most frequently were in the areas of statistics and methodology. No one can be a competent contemporary psychologist without being fully conversant with statistics and probability.

Clearly, one of the goals of this book is to make research in the discipline of psychology more accessible to the general reader. However, the empirical methods and techniques of theory construction in psychology are so intertwined with statistics (as is the case in many other fields, such as economics, sociology, and genetics) that it would be wrong to imply that one can thoroughly understand the field without having some statistical knowledge. Thus, although this chapter has served as an extremely sketchy lesson in statistical thinking, its main purpose has been to highlight the existence of an area of expertise that is critical to a full understanding of psychology.

SUMMARY

As in most sciences, the conclusions that are drawn from psychological research are probabilistic conclusions—generalizations that hold more often than not but that do not apply in every single case. The predictions derived from psychological findings and theories are still useful even though they are not 100 percent accurate (just as is the case in virtually all medical treatments, for example, which are effective only in a probabilistic sense).

One thing that prevents the understanding of much psychological research is that many people have difficulty thinking in probabilistic terms. In this chapter, we discussed several well-researched examples of how probabilistic reasoning goes astray for many people: They make insufficient use of probabilistic information when they also have vivid testimonial evidence available. They invert conditional probabilities, thus acting as if the probability of A given B is the same as the probability of B given A, which it is not. They fail to take into account the fact that larger samples give more accurate estimates of population values. And finally, they display the gambler's fallacy: the tendency to see links among events that are really independent. This fallacy derives from a more general tendency that we will discuss in the next chapter: the tendency to fail to recognize the role of chance in determining outcomes.