CHAPTER 5

CORRELATION AND CAUSATION

BIRTH CONTROL BY THE TOASTER METHOD

Several years ago, a large-scale study of the factors related to the use of contraceptive devices was conducted in Taiwan. A large research team of social scientists and physicians collected data on a wide range of behavioral and environmental variables. The researchers were interested in seeing what variables best predicted the adoption of birth control methods. After collecting the data, they found that the one variable most strongly related to contraceptive use was the number of electrical appliances (toaster, fans, etc.) in the home (Li, 1975).

This result probably does not tempt you to propose that the teenage pregnancy problem should be dealt with by passing out free toasters in high schools. But why aren't you tempted to think so? The correlation between appliances and contraceptive use was indeed strong, and this variable was the single best predictor among the many variables that were measured. Your reply, I hope, will be that it is not the strength but the *nature* of the relationship that is relevant. Starting a free toaster program would imply the belief that toasters *cause* people to use contraceptives. The fact that we view this suggestion as absurd means that, at least in clear-cut cases such as this, we recognize that two variables may be associated without having a causal relationship.

In this example, we can guess that the relationship exists because contraceptive use and the number of electrical appliances in the home are linked through some other variable that relates to both. Education would be one likely candidate for a mediating variable. We know that educational level is related to both contraceptive use and socioeconomic status. All we need now is the fact that families at higher socioeconomic levels tend to have more electrical appliances in their homes, and we have the linkage. Of course, other variables may mediate this correlation. However, the point is that, no matter how strong the correlation is between the number of toasters and contraceptive use, the relationship does not indicate a causal connection.

The contraceptive example makes it very easy to understand the fundamental principle of this chapter: The presence of a correlation does not necessarily imply causation. However, the limitations of correlational evidence are not always so easy to recognize. When the causal link seems obvious to us, when we have a strong preexisting bias, or when our interpretations become dominated by our theoretical orientation, it is tempting to treat correlations as evidence of causation. The case of Karl Pearson provides a clear and ironic example.

PEARSON AND TUBERCULOSIS

Pearson was a British statistician of considerable eminence who lived at the turn of the century. He developed a mathematical formula to express the strength of the relationship between two variables. Scientists cannot be satisfied with the vague statement that two variables are correlated; they must have some way of measuring the strength of the correlation. Pearson provided one of the most useful formulas to calculate this strength, and this "correlation coefficient" now bears his name. However, the father of the correlation coefficient set a disastrous precedent, for he was one of the most blatant violators of the principle that correlation does not necessarily imply causation.

Pearson calculated the correlations between many biological variables. One of his findings was that the correlation between the tendency of a parent to contract tuberculosis and the tendency of the parent's children to contract the same disease was .50. (Correlation coefficients range from 0 to 1.0 in absolute value; higher numbers indicate stronger relationships; negative values indicate inverse relationships.) Now, for Pearson this was far from a neutral finding, for he was an extreme hereditarian, a leader in the growing eugenics movement of his day. Pearson thought that virtually every imaginable human characteristic was primarily a product of genetic influences and used this belief to justify his prejudices. He argued, for example, that Jews living in slums were "innately dirty" (see Blum, 1978). Thus, he saw the parent-child tuberculosis correlation as evidence of a hereditary cause of the disease. He dismissed environmental explanations and argued against sanitation campaigns on the grounds that the hereditary nature of the disease made them futile.

Of course, the discovery of the tubercle bacillus and the reduction of the disease by improved sanitary conditions have rendered a harsh historical judgment on Pearson's views. His example illustrates the difficulty of refraining from causal interpretations of correlational data when a preexisting bias is present. This example also points to the importance of maintaining the strictest standards for evaluating evidence when the issue at hand is one of pressing social importance. Incorrect beliefs have serious consequences.

Those who endorse the attitude that anything goes in the field of psychology should take heed. Many of our current pressing social problems come within the domain of psychology. The abandonment or distortion of the empirical attitude toward social problems can have disastrous effects, as an examination of the history of hunger in America illustrates.

THE THIRD-VARIABLE PROBLEM: GOLDBERGER AND PELLAGRA

In the early 1900s, thousands of Americans in the South suffered and died (approximately 100,000 fatalities per year) of a disease called *pellagra.* Characterized by dizziness, lethargy, running sores, vomiting, and severe diarrhea, the disease was thought to be infectious and to be caused by a living microorganism of "unknown origin" (Chase, 1977, p. 205). It is not surprising, then, that many physicians of the National Association for the Study of Pellagra were impressed by evidence that the disease was linked to sanitary conditions. It seemed that homes in Spartanburg, South Carolina, that were free of pellagra invariably had inside plumbing and good sewerage. This correlation coincided quite well with the idea of an infectious disease transmitted, because of poor sanitary conditions, via the excrement of pellagra victims.

One physician who doubted this interpretation was Joseph Goldberger, who, at the direction of the surgeon general of the United States, had conducted several investigations of pellagra. Goldberger thought that pellagra was caused by inadequate diet—in short, by the poverty common throughout the South. Many victims had lived on high-carbohydrate, extremely low-protein diets, characterized by small amounts of meat, eggs, and milk and large amounts of corn, grits, and mush. Goldberger thought that the correlation between sewage conditions and pellagra did not reflect a causal relationship in either direction (much as in the toaster–birth-control example). Goldberger thought that the correlation arose because families with sanitary plumbing were likely to be economically advantaged. This economic discrepancy would also be reflected in their diets, which would contain more animal protein.

But wait a minute! Why should Goldberger get away with *his* causal inference? After all, both sides were just sitting there with their correlations, Goldberger with pellagra and diet and the other physicians with pellagra and sanitation. Why shouldn't the association's physicians be able to say that Goldberger's correlation was equally misleading? Why was he justified in rejecting the hypothesis that an infectious organism was transmitted through the excrement of pellagra victims because of inadequate sewage disposal? Well, the reason Goldberger was justified has to do with one small detail that I neglected to mention: Goldberger had eaten the excrement of pellagra victims.

Why Goldberger's Evidence Was Better

Goldberger had a type of evidence (a controlled manipulation, discussed further in the next chapter) that is derived when the investigator, instead of simply observing correlations, actually manipulates the critical variable. This approach often involves setting up special conditions that rarely occur naturally, and to term Goldberger's special conditions unnatural is an understatement.

Confident that pellagra was not contagious and not transmitted by the bodily fluids of the victims, Goldberger had himself injected with the blood of a victim. He inserted throat and nose secretions from a victim into his own mouth. In addition:

> Finally, he selected two patients—one with scaling sores and the other with diarrhea. He scraped the scales from the sores, mixed the scales with four cubic centimeters of urine from the same patients, added an equal amount of liquid feces, and rolled the mixture into little dough balls by the addition of four pinches of flour. The pills were taken voluntarily by him, by his assistants and by his wife. (Bronfenbrenner & Mahoney, 1975, p. 11)

Neither Goldberger nor the other volunteers came down with pellagra. In short, Goldberger had created the conditions necessary for the infectious transmission of the disease, and nothing had happened.

Goldberger had now manipulated the causal mechanism suggested by others and had shown that it was ineffective, but it was still necessary to test his own causal mechanism. Goldberger got two groups of prisoners from a Mississippi state prison farm who were free of pellagra to volunteer for his experiment. One group was given the high-carbohydrate, low-protein diet that he suspected was the cause of pellagra, while the other group received a more balanced diet. Within five months, the low–protein group was ravaged by pellagra, while the other group showed no signs of the disease. After a long struggle, during which Goldberger's ideas were opposed by those with political motives for denying the existence of poverty, his hypothesis was eventually accepted because it matched the empirical evidence better than any other.

The history of pellagra illustrates the human cost of basing social and economic policy on mistaken inferences from correlational studies. This is not to say that we should never use correlational evidence. Quite the contrary. In many instances, it is all we have to work with (see Chapter 8), and in some cases, it is all we need (for instance, when prediction, rather than determination of cause, is the goal). Scientists often have to use incomplete knowledge to solve problems. The important thing is that we approach correlational evidence with a certain skepticism. Examples such as the pellagra–sewage case occur with considerable frequency in all areas of psychology. The example illustrates what is sometimes termed the *third-variable problem:* the fact that the correlation between the two variables—in this case, pellagra incidence and sewage conditions—may not indicate a direct causal path between them but may arise because both variables are related to a third variable—here, diet—that has not even been measured. Correlations like that between sewage and pellagra are often termed *spurious correlations:* correlations that arise not because a causal link exists between the two variables that are measured, but because both variables are related to a third variable.

Let's consider a more contemporary example. During the 1980s, debates raged over the relative efficacy of public and private schools. Some of the conclusions drawn in this debate vividly demonstrate the perils of inferring causation from correlational evidence. The question of the efficacy of private versus public schools is an empirical problem that can be attacked with the investigative methods of the social sciences. This is not to imply that it is an easy problem, only that it is a scientific problem, potentially solvable. All advocates of the superiority of private schools implicitly recognize this, because at the crux of their arguments is an empirical fact: Student achievement in private schools exceeds that in public schools. This fact is not in dispute—educational statistics are plentiful and largely consistent across various studies. The problem is the use of these achievement data to conclude that the education received in private schools *causes* the superior test scores.

The outcome of educational testing is a function of many different variables, all of which are correlated. In order to evaluate the relative efficacy of public schools and private schools, we need more complex statistics than merely the relationship between the type of school attended and school achievement. For example, educational achievement is related to many different indicators of family background, such as parental education, parental occupation, socioeconomic status, the number of books in the home, and other factors. These characteristics are also related to the probability of sending a child to a private school. Thus, family background is a potential third variable that may affect the relationship between academic achievement and the type of school. In short, the relationship may have nothing to do with the effectiveness of private schools but may be the result of the fact that economically advantaged children do better academically and are more likely to attend private schools.

Fortunately there exist complex correlational statistics known as *multiple regression* and *path analysis* (statistics developed in part by psychologists; see Chapter 12) that were designed to deal with problems such as this one. These statistics allow the correlation between two variables to be recalculated after the influence of other variables is removed, or "factored out" or "partialed out." Using these more complex correlational techniques, Ellis Page and Timothy Keith (1981), two researchers at Duke University, analyzed a large set of educational statistics on high-school students that were collected under the auspices of the National Center for Educational Statistics. They found that, after variables reflecting the students' home backgrounds and general mental ability were factored out, there was virtually no relationship between school achievement and the type of school attended.

Their results have been confirmed by other researchers (Berliner & Biddle, 1995; Jencks, 1985; Walberg & Shanahan, 1983; Wolfle, 1987).

Thus, it appears that advocating private schools as a means of improving educational achievement is the same as arguing for birth control by the toaster method. Academic achievement is linked to private school attendance not because of any direct causal mechanism, but because the family background and the general cognitive level of students in private schools are different from those of children in public schools.

The complex correlational statistics that allow us to partial out the effects of a third variable do not always reduce the magnitude of the original correlation. Sometimes the original correlation between two variables remains even after the partialing out of the third variable, and this result itself can be informative. Such an outcome indicates that the original correlation was not due to a spurious relationship with that particular third variable. Of course, it does not remove the possibility of a spurious relationship due to some other variable.

A good example is provided in the data analyzed by Thomas, Alexander, and Eckland (1979). These investigators found that the probability that a high-school student will attend university is related to the socioeconomic status of the student's family. This is an important finding that strikes at the heart of the merit-based goals of our society. It suggests that opportunities for success in life are determined by a person's economic class. However, before jumping to this conclusion, we must consider several other alternative hypotheses. That is, the correlation between university attendance and socioeconomic status should be examined closely for spuriousness. One obvious candidate for a third variable is academic ability. Perhaps this is related to both university attendance and socioeconomic status, and if it is partialed out, the correlation between the first two variables may disappear. The investigators calculated the appropriate statistics and found that the correlation between university attendance and socioeconomic status remained significant even after academic aptitude was partialed out. Thus, the fact that children of higher economic classes are more likely to attend university is not entirely due to differences in academic aptitude. This finding, of course, does not rule out the possibility that some other variable leads to the relationship between the first two, but it is clearly important, both practically and theoretically, to be able to rule out a major alternative explanation such as academic aptitude.

Anderson and Anderson (1996) describe how they tested a theory of regional differences in violence by examining whether a set of competing hypotheses could account for the data instead. They used the techniques of partial correlation to do this. It turns out that the violent crime in the United States is higher in the southern states than in the northern states. Anderson and Anderson (1996) tested what has been called the *heat hypothesis*—that "uncomfortably warm temperatures produce increases in aggressive motives and (sometimes) aggressive behavior" (p. 740). Not surprisingly, they did find a correlation between the average temperature in a city and its violent crime rate. What gives the heat hypothesis more credence, however, is that they found that the correlation between temperature and violent crime remained significant even after variables such as unemployment rate, per capita income, poverty rate, education, population size, median age of population, and several other variables were statistically controlled.

Finally, recall the example of the research on the effects of teenagers' work experience described in Chapter 1. Developmental psychologists have found that more work experience is associated with poorer school achievement and higher delinquency (Bachman & Schulenberg, 1993; Greenberger & Steinberg, 1986; Steinberg, Brown, & Dornbusch, 1996; Steinberg, Fegley, & Dornbusch, 1993). However, this evidence, by itself, is merely correlational. It may be the case, for example, that teenagers who choose to work a lot have lower academic ability and/or different home backgrounds to begin with. Researchers have tested these possible third variables with the complex correlational statistics previously mentioned and have found that the negative associations still remain after academic ability and other background variables have been controlled (Bachman & Schulenberg, 1993; Steinberg, Fegley, & Dornbusch, 1993).

THE DIRECTIONALITY PROBLEM

There is no excuse for making causal inferences on the basis of correlational evidence when it is possible to manipulate variables in a way that would legitimately justify a causal inference. Yet this is a distressingly common occurrence when psychological issues are involved, and the growing importance of psychological knowledge in the solution of social problems is making this tendency increasingly costly. A well-known example in the area of educational psychology illustrates this point quite well.

Since the beginning of the scientific study of reading about a hundred years ago, researchers have known that there is a correlation between eye movement patterns and reading ability. Poorer readers make more erratic movements, display more regressions (movements from right to left), and make more fixations (stops) per line of text. On the basis of this correlation, some educators hypothesized that deficient oculomotor skills were the cause of reading problems, and many eye-movement-training programs were developed and administered to elementary-school children. These programs were instituted long before it was ascertained whether the correlation really indicated that erratic eye movements caused poor reading.

It is now known that the eye-movement–reading-ability correlation reflects a causal relationship that runs in exactly the opposite direction. Erratic eye movements do not cause reading problems (Olson & Forsberg, 1993). Instead, slow recognition of words and difficulties with comprehension lead to erratic eye movements. When children are taught to recognize words efficiently and to comprehend better, their eye movements become smoother.

Training children's eye movements does nothing to improve their reading comprehension.

Since the mid-1970s, research has clearly pointed to word decoding and a language problem in phonological processing as the sources of reading problems (Share, 1995; Shaywitz, 1996; Stanovich, 2000; Stanovich & Siegel, 1994). Very few cases of reading disability are due to difficulties in the area of eye movement patterns (Olson & Forsberg, 1993; Share & Stanovich, 1995). Yet, if most school districts of at least medium size were to search diligently in their storage basements, they would find dusty eye movement trainers that represent thousands of dollars of equipment money wasted because of the temptation to see a correlation as proof of a causal hypothesis.

Another somewhat similar example is provided by the enthusiasm for self-esteem as an explanatory construct in the 1990s. An extremely popular hypothesis was that school achievement problems, drug abuse, teenage pregnancy, and many other problem behaviors were the result of low self-esteem (Dawes, 1994; Kahne, 1996). It was assumed that the causal direction of the linkage was obvious: Low self-esteem led to problem behaviors, and high self-esteem led to high educational achievement and accomplishments in other domains. This assumption of causal direction provided the motivation for many educational programs for improving self-esteem. The problem here was the same as that in the eye movement example: An assumption of causal direction was made from the mere existence of a correlation (in this case, quite a small correlation; see Kahne, 1996). It turns out that the relationship between self-esteem and school achievement, if it exists at all, is just as likely to be in the opposite direction: Superior accomplishment in school (and in other aspects of life) leads to high self-esteem (Dawes, 1994; Kahne, 1996).

Our discussion thus far has identified the two major classes of ambiguity present in a simple correlation between two variables. One is called the *directionality problem* and is illustrated by the eye movement and self-esteem examples. Before immediately concluding that a correlation between variable A and variable B is due to changes in A causing changes in B, we must first recognize that the direction of causation may be the opposite, that is, from B to A. The second problem is the third-variable problem, and it is illustrated by the pellagra example (and the toaster–birth-control and private-school–achievement examples). The correlation between the two variables may not indicate a causal path in either direction but may arise because both variables are related to a third variable.

SELECTION BIAS

There are certain situations in which the possibility of a spurious correlation is very likely. These are situations in which there is a high probability that selection bias has occurred. The term *selection bias* refers to the relationships between certain subject and environmental variables that may arise when people with different biological, behavioral, and psychological characteristics select different types of environments. Selection bias creates a spurious correlation between environmental characteristics and behavioral-biological characteristics.

Let's look at a straightforward example that illustrates the importance of selection factors in creating spurious correlations: Quickly name the state with the highest incidence of deaths due to respiratory illness. The answer to this question is, of course, Arizona. What? Wait a minute! Arizona has clean air, doesn't it? Does the smog of Los Angeles spread that far? Has the suburban sprawl of Phoenix become that bad? No, it can't be. Let's slow down a minute. Maybe Arizona *does* have good air. And maybe people with respiratory illnesses tend to move there. And then they die. There you have it. A situation has arisen in which, if we're not careful, we may be led to think that Arizona's air is killing people.

However, selection factors are not always so easy to discern. They are often overlooked, particularly when, as in Pearson's case, there is a preexisting desire to see a certain type of causal link. Tempting correlational evidence combined with a preexisting bias may deceive even the best of minds. Let's consider some specific cases.

The importance of considering selection factors was illustrated quite well in the national debate over the quality of American education that took place throughout most of the 1980s and continued into the 1990s. During this debate, the public was inundated with educational statistics but was not provided with corresponding guidance for avoiding the danger of inferring causal relationships from correlational data that are filled with misleading selection factors.

Throughout the continuing debate, many commentators with a political agenda repeatedly attempted to provide evidence that educational quality is not linked to teacher salary levels or class size, despite evidence that both are important (Finn & Achilles, 1999). One set of findings put forth was the Scholastic Aptitude Test (SAT) results for each of the 50 states. The average scores on this test, taken by high-school students who intend to go to certain universities, did indeed show little relation to teacher salaries and general expenditure on education. If anything, the trends seemed to run opposite to the expected direction. Several states that had very high average teacher salaries had very low average SAT scores, and many states at the bottom of the teacher salary rankings had very high average SAT scores. A close look at the data patterns provides an excellent lesson in how easily selection factors can produce spurious correlations.

On further examination, we see that Mississippi students, for example, score higher than California students on the SAT (Powell & Steelman, 1996; Taube & Linden, 1989). In fact, the difference is considerable. The average Mississippi scores are over 100 points higher. Because Mississippi teachers' salaries are the lowest in the nation, this was cause for celebration among conservative commentators arguing for cuts in teachers' salaries. But wait. Is it really true that schools are better in Mississippi than in California, that the general state of education is superior in the former? Of course not. Virtually any other objective index would show that California schools are superior (Powell & Steelman, 1996). But if this is true, what about the SAT?

The answer lies in selection factors. The SAT is not taken by *all* high-school students. Unlike much standardized testing that schools conduct, in which all children are uniformly tested, the SAT involves selection bias (Powell & Steelman, 1996; Taube & Linden, 1989; Wainer, 1989). Only students hoping to go to a university take the test. This factor accounts for some of the state-by-state variance in average scores on the test and also explains why some of the states with the very best educational systems have very low average SAT scores.

Selection factors operate on the SAT scores of states in two different ways. First, some state university systems require the American College Testing (ACT) program test scores rather than the SAT scores. Thus, the only students who take the SAT in these states are students planning to go to a university out of state. It is more likely that these students will be from advantaged backgrounds and/or will have higher academic aptitude than the average student. This is what happened in the Mississippi-California example. Only 4 percent of Mississippi high-school students took the SAT, whereas the figure in California was 47 percent (Powell & Steelman, 1996).

The second selection factor is a bit more subtle. In states with good educational systems, many students intend to continue their education after high school. In such states, a high proportion of students take the SAT, including a greater number with lesser abilities. States with high dropout rates and lower overall quality have a much smaller proportion of students who aspire to university. The group of students who eventually take the SAT in such states represents only those best qualified to go to a university. The resulting average SAT scores in these states naturally tend to be higher than those from states where larger proportions of students pursue further education.

The misuse of SAT scores also provides us with an unfortunate example of how hard it is to correct the misleading use of statistics as long as the general public lacks the simple methodological and statistical thinking skills taught in this book. I included an example of the misuse of SAT scores due to selection biases in the first edition of this book, written in 1983. In the fourth edition, written over ten years later in 1994, I discussed an article by Brian Powell (1993), an Indiana professor, analyzing a column written by political columnist George Will in 1993 in which Will—you guessed it—argued against public expenditures on education because states with high SATs do not have high expenditures on education. Powell (1993) pointed out that the states that Will singled out as having particularly high scores—Iowa, North Dakota, South Dakota, Utah, and Minnesota—have SAT participation rates of only 5 percent, 6 percent, 7 percent, 4 percent, and 10 percent, respectively, whereas more than 40 percent of all high-school seniors in the United States take the SAT. The reason is that in these states, the test required for admission to public institutions is the ACT test. Only students planning on studying out of state, "often at prestigious private schools" (Powell, 1993, p. 352), take the SAT. In contrast, in New Jersey, which Will used as an example of a state with low SAT scores and high expenditures, 76 percent of high-school seniors take the test. Obviously the students in North and South Dakota who take the SAT are a more select group than those in New Jersey, where three-quarters of all students take the test.

In the journal *Educational Researcher,* psychometrician Howard Wainer (1993) analyzed an article in the June 22, 1993, *Wall Street Journal* that featured a study by the Heritage Foundation, an ideologically biased think tank. The foundation argued against spending money on education because—you guessed it—SAT scores were lower in states where more money was spent on education. Wainer's article goes beyond merely showing how selection bias causes this result, however. Wainer demonstrated that when a test that uses a representative sample rather than a self-selected sample is analyzed (the National Assessment for Educational Progress, or NAEP), the relationship reverses: States that spend more on education have higher scores.

Using the partial correlation techniques mentioned earlier, Powell and Steelman (1996) confirmed this relationship. They found that once the states were statistically equated for the proportion of students who took the test, each additional $1,000 of per pupil expenditure was associated with a 15-point *increase* in the average SAT scores for the state. Nevertheless, despite the overwhelming evidence that selection effects make the state-by-state comparison of SAT scores meaningless unless statistical adjustments are made, the media and politicians continue to use the unadjusted scores to advance political agendas (in 1994, a candidate for the governor of New York state "cited the state's SAT rank as reason to oust the incumbent"; Powell & Steelman, 1996, p. 31). Readers are invited to help me correct this misuse of SAT scores in the media. Write to the media outlet that you detect using SAT scores to compare states, and inform them of how they are misusing statistics. Please send me any responses that you receive.

An example from clinical psychology demonstrates how tricky and "perverse" the selection bias problem can be. It has sometimes been demonstrated that the cure rate for various addictive-appetite problems such as obesity, heroin use, and cigarette smoking is *lower* for those who have had psychotherapy than for those who have not (Rzewnicki & Forgays, 1987; Schachter, 1982). The reason, you will be glad to know, is not that psychotherapy makes addictive behavior more resistant to change. It is that, among those who seek psychotherapy, the disorder is more intractable, and self-cures have been ineffective.

Wainer (1999) tells a story from World War II that reminds us of the sometimes perverse aspects of selection bias. He describes an aircraft analyst who was trying to determine where to place extra armor on an aircraft based on the pattern of bullet holes in the returning planes. His decision was to put the extra armor in the places that were *free* of bullet holes on the returning aircraft that he analyzed. His reasoning was that the planes had probably been pretty uniformly hit with bullets. Where he found the bullet holes on the returning aircraft told him that, in those places, the plane could be hit and still return. Those areas that were free of bullet holes on returning planes had probably been hit—but planes hit there did not return. Hence the places on the returning planes without bullet holes needed more armor.

In short, the consumer's rule for this chapter is simple: Be on the lookout for instances of selection bias, and avoid inferring causation when data are only correlational. It is true that complex correlational designs do exist that allow limited causal inferences, and that some research problems are structured in such a way that correlational evidence allows one to choose between causal hypotheses. It is also true that correlational evidence is helpful in demonstrating convergence on a hypothesis (see Chapter 8). Nevertheless, it is probably better for the consumer to err on the side of skepticism than to be deceived by correlational relationships that falsely imply causation.

SUMMARY

The central point of this chapter was to convey that the mere existence of a relationship between two variables does not guarantee that changes in one are causing changes in the other—that correlation does not imply causation. Two problems in interpreting correlational relationships were discussed. In the third-variable problem, the correlation between the two variables may not indicate a direct causal path between them but may arise because both variables are related to a third variable that has not even been measured. If, in fact, the potential third variable has been measured, correlational statistics such as partial correlation (to be discussed again in Chapter 8) can be used to assess whether that third variable is determining the relationship. The other thing that makes the interpretation of correlations difficult is the existence of the directionality problem: the fact that even if two variables have a direct causal relationship, the direction of that relationship is not indicated by the mere presence of the correlation.

Selection bias is the reason for many spurious relationships in the behavioral sciences: the fact that people choose their own environments to some extent and thus create correlations between behavioral characteristics and environmental variables. As the example of Goldberger illustrated, and as will be illustrated extensively in the next two chapters, the only way to ensure that selection bias is not operating is to conduct a true experiment in which the variables are manipulated.