Research Methods**: Establishing the Facts**

Psychologists gather evidence to support their hypotheses by using different methods, depending on the kinds of questions they want to answer. These methods are not mutually exclusive, however. Just as a police detective may use a magnifying glass *and* a fingerprint duster *and* interviews of suspects to figure out "who done it," psychological sleuths often draw on different techniques at different stages of an ongoing investigation. As you read about these methods, you may want to list their advantages and disadvantages so you will remember them better; then check your list against the one in Figure 1.

We will begin with descriptive methods, which allow researchers to describe and predict behavior but not necessarily to choose one explanation over competing ones.

Descriptive Methods

Case Studies

A case study (or *case history)* is a detailed description of a particular individual, based on careful observation or formal psychological testing. It may include information about a person's childhood, dreams, fantasies, experiences, relationships, and hopes--anything that will provide insight into the person's behavior. Case studies are most commonly used by clinicians, but sometimes academic researchers use them as well, especially when they are just beginning to study a topic or when practical or ethical considerations prevent them from gathering information in other ways.

For example, suppose you want to know whether the first few years of life are critical for acquiring a first language. Can children who have missed out on hearing speech (or, in the case of deaf children, seeing signs) "catch up" later? Obviously, psychologists cannot answer this question by isolating children and seeing what happens! So instead they have studied unusual cases of language deprivation.

One such case involved a 13-year-old girl who had been cruelly locked up in a small room since the age of 1 1/2, strapped for hours to a potty chair. Her mother, a battered wife, barely cared for her, and although the child may have heard some speech through the walls of her room, no one in the family spoke a word to her. If she made the slightest sound, her severely disturbed father beat her with a large piece of wood. When she was finally rescued, "Genie," as researchers called her, did not know how to chew or stand erect, was not toilet trained, and spat on anything that was handy, including other people. Her only sounds were high-pitched whimpers. Eventually, she was able to learn some rules of social conduct, and she began to understand short sentences and to use words to convey her needs, describe her moods, and even lie. Yet even after many years, Genie's grammar and pronunciation remained abnormal. She never learned to use pronouns correctly, ask questions, produce proper negative sentences, or use the little word endings that communicate tense, number, and possession (Curtiss, 1977, 1982; Rymer, 1993). This sad case, along with similar ones, suggests that a critical period exists for language development--a window of opportunity for acquiring language--with the likelihood of fully mastering a first language declining steadily after early childhood and falling off drastically at puberty (Pinker, 1994).

Case studies illustrate psychological principles in a way that abstract generalizations and cold statistics never can, and they produce a more detailed picture of an individual than other methods do. In biological research, cases of patients with brain damage have yielded important clues to how the brain is organized. But in most instances, case studies have serious drawbacks. Information is often missing, or is hard to interpret; for example, no one knows what Genie's language development was like before she was locked up, or whether she was born with mental deficits. The observer who writes up the case is bound to have certain biases that influence which facts he or she notices and includes--or omits.

And the subject of the case study may have selective and inaccurate memories of his or her past experiences, making the observer's conclusions somewhat unreliable. Most important, this method has limited usefulness for deriving general principles of behavior because the person who is the focus of the case study may be *unrepresentative* of the group that a researcher is interested in. For all these reasons, case studies are usually only sources, rather than tests, of hypotheses.

When people do draw conclusions solely on the basis of case studies, the results can be disastrous. For example, many clinicians once believed that autism in children was caused by rejecting, cold, "refrigerator" mothers. This belief was based on the writings of psychoanalyst Bruno Bettelheim (1967), who drew his conclusions from three published cases of autistic children whose mothers had psychological problems, and from a few other unpublished cases whose number he exaggerated (Pollak, 1997). When proper studies were finally done, using objective testing procedures and a larger, representative group of autistic children and their parents, scientists learned that parents of autistic children are as psychologically healthy as any other parents. Today we know that autism stems from a neurological problem rather than from any psychological problems of the mothers, and that certain genes may increase susceptibility to the disorder

(Ingram et aI., 2000). But because so many people accepted Bettelheim's claims, thousands of women blamed themselves for their children's disorder and suffered needless guilt and remorse.

Be wary, then, of the compelling cases reported in the media by individuals ("I was a multiple personality") or therapists. Often, these stories are only "arguing by anecdote," and they are not a basis for drawing firm conclusions about anything.

Observational Studies

In observational studies, the researcher observes, measures, and records behavior, taking care to avoid intruding on the people (or animals) being observed. Unlike case studies, observational studies usually involve many participants ("subjects"). Often an observational study is the first step in a program of research; it is helpful to have a good description of behavior before you try to explain it.

The primary purpose of *naturalistic observation* is to find out how people or animals act in their normal social environments. Ethologists such as Jane Goodall and the late Dian Fossey used this method to study apes and other animals in the wild. Psychologists use naturalistic observation wherever people happen to be--at home, on playgrounds or streets, in schoolrooms, or in offices. In one study, a social psychologist and his students ventured into a common human habitat: bars. They wanted to know whether people in bars drink more when they are in groups than when they are alone. They visited all 32 pubs in a mid-sized city, ordered beers, and recorded on napkins and pieces of newspaper how much the other patrons imbibed. They found that drinkers in groups consumed more than individuals who were alone. Those in groups did not drink any faster; they just lingered in the bar longer (Sommer, 1977).

Note that the students who did this study did not rely on their impressions or memories of how much people drank. In observational studies, researchers count, rate, or measure behavior in a systematic way. These procedures help to minimize the tendency of observers to notice only what they expect or want to see. Careful record keeping ensures accuracy and allows different observers to cross-check their observations for consistency. Observers must also take pains to avoid being obvious about what they are doing and to disguise their intentions so they can see people as they really are. If the researchers who studied drinking habits had marched in with video cameras and announced that they were psychology students, the bar patrons might not have behaved naturally.

Sometimes psychologists prefer to make observations in a laboratory setting. In *laboratory observation,* psychologists have more control of the situation. They can use sophisticated equipment, maintain a clear line of vision, remain hidden behind a one-way mirror, and so forth. Suppose that you wanted to know how infants of different ages respond when left with a stranger. The most efficient approach might be to have parents and their infants come to your laboratory, observe them playing together for a while through a one-way window, then have a stranger enter the room and, a few minutes later, have the parent leave. You could record signs of distress, interactions with the stranger, and other behavior. If you did this, you would find that very young infants carry on cheerfully with whatever they are doing when the parent leaves. However, by the age of about 8 months, many children will burst into tears or show other signs of what child psychologists call "separation anxiety."

One shortcoming of laboratory observation is that the presence of researchers and special equipment may cause subjects to behave differently than they would in their usual surroundings. Further, observational studies, like other descriptive studies, are more useful for describing behavior than for explaining it. For example, the barroom results we described do not necessarily mean that being in a group makes people drink a lot. People may join a group because they are already interested in drinking and find it more comfortable to hang around the bar if they are with others. Similarly, if we observe infants protesting whenever a parent leaves the room, we cannot be sure *why* they are protesting. Is it because they have become attached to their parents and want them nearby? Is it because they have learned from experience that crying brings an adult with a cookie and a cuddle? Observational studies alone cannot answer such questions.

Surveys

Psychological tests usually generate information about people indirectly. In contrast, surveys are questionnaires and interviews that gather information by asking people *directly* about their experiences, attitudes, or opinions. Most of us are familiar with national opinion surveys, such as the Gallup and Roper polls. Surveys have been done on hundreds of topics, from consumer preferences to sexual preferences.

Surveys produce bushels of data, but they are not easy to do well. The biggest hurdle is getting a representative sample, a group of subjects that accurately represents the larger population that the researcher wishes to describe. Suppose you wanted to know about drug use among college sophomores. Questioning every sophomore in the country would obviously not be practical; instead, you would need to recruit a sample. You could use special selection procedures to ensure that this sample contained the same proportion of women, men, blacks, whites, poor people, rich people, Catholics, Jews, and so on as in the general population of college sophomores. Even then, a sample drawn just from your own school or town might not yield results applicable to the entire country, or even state.

Most people do not realize that a sample's size is less critical than its representativeness. A small but representative sample may yield extremely accurate results, whereas a surveyor poll that fails to use proper sampling methods may yield questionable results, no matter how large the sample. For example, when a talk-radio host surveys listeners about a political issue, or a magazine surveys its readers about their sexual habits, the results are not likely to generalize to the population as a whole--even if thousands of people respond. Why? As a group, people who listen to talk radio or read *Cosmo* are likely to hold different opinions than those who prefer, say, a classical music station or read *Scientific American.*

Popular polls and surveys also suffer from a volunteer bias. People who feel strongly enough to volunteer their opinions may differ from those who remain silent. When you read about a survey (or any other kind of study), always ask who participated. A biased, nonrepresentative sample does not necessarily mean that a survey is worthless or uninteresting, but it does mean that the results may not hold true for other groups.

Another problem with surveys is that people sometimes lie--especially when the survey is about a touchy topic ("What? Me do that disgusting/illegal/dishonest thing? Never!"). The likelihood of lying is reduced when respondents are guaranteed anonymity. Computer technology can help in this regard, because many people feel more anonymous when they "talk" to a computer than when they fill out a paper-and-pencil questionnaire. In one national study of HIV-risk sexual behaviors, violence, and drug use, teenage boys who responded on a computer keyboard to digitally recorded questions played through headphones, were far more likely to admit to risky behaviors than were boys who filled out questionnaires (Turner et al, 1998). Researchers also have ways to check for lying--for example, by asking the same question several times with different wording and checking for consistency in the answers. But not all surveys use these techniques, and even when respondents are trying to be truthful, they may misinterpret the survey questions or misremember the past.

Finally, when you hear about the results of a surveyor opinion poll, you need to consider which questions were (and were not) asked and how the questions were phrased. The questions a researcher asks may reflect his or her assumptions about the topic or may be designed to further a particular agenda (Ericksen & Steffen, 1999). And the phrasing of a question can affect how people respond to it (as political pollsters know well). Many years ago, famed sex researcher Alfred Kinsey, in his pioneering surveys of sexual habits (Kinsey, Pomeroy, & Martin, 1948; Kinsey et al, 1953), made it his practice always to ask, *"How many times have you* (masturbated, had non-marital sex, etc.)?" rather than *"Have you ever* (masturbated, had nonmarital sex, etc.)?" The first way of phrasing the question tended to elicit more truthful responses than the second, because it removed the respondent's potential self-consciousness about having done any of these things. The second way of phrasing the question would have permitted embarrassed respondents to reply with a simple but dishonest "No."

As you can see, although surveys can be extremely informative, they must be conducted and interpreted carefully.

Correlational Studies: Looking for Relationships

Psychologists often want to know whether two or more phenomena are related and, if so, how strongly. For example, are students' grade-point averages related to the number of hours they spend watching television? To find out, a psychologist would do a correlational study.

Measuring Correlations

The word correlation is often used as a synonym for relationship. Technically, however, a correlation is a numerical measure of the *strength* of the relationship between two things. The "things" may be events, scores, or anything else that can be recorded and tallied. In psychological studies, such things are called variables because they can vary in quantifiable ways. Height, weight, age, income, IQ scores, number of items recalled on a memory test, number of smiles in a given time period--anything that can be measured, rated, or scored can serve as a variable. I

Correlations always occur between *sets* of observations. In psychological research, these sets of observations usually come from many individuals or are used to compare groups of people. For example, in research on the origins of intelligence, psychologists look for a correlation between the IQ scores of parents and those of their children. To do this, the researchers must gather scores from a set of parents and from the children of these parents. You cannot compute a correlation if you know the IQs of only one particular parent-child pair. To say that a relationship exists, you need more than one pair of values to compare.

A positive correlation means that high values of one variable are associated with high values of the other, and that low values of one variable are associated with low values of the other. Height and weight are positively correlated, for example; so are IQ scores and school grades. Rarely is a correlation perfect, however. Some tall people weigh less than some short ones; some people with average IQs are superstars in the classroom and some with high IQs get poor grades. Figure 2.2 shows a positive correlation between men's educational level and their annual income. Each dot represents a man; you can find each man's educational level by drawing a horizontal line from his dot to the vertical axis. You can find his income by drawing a vertical line from his dot to the horizontal axis.

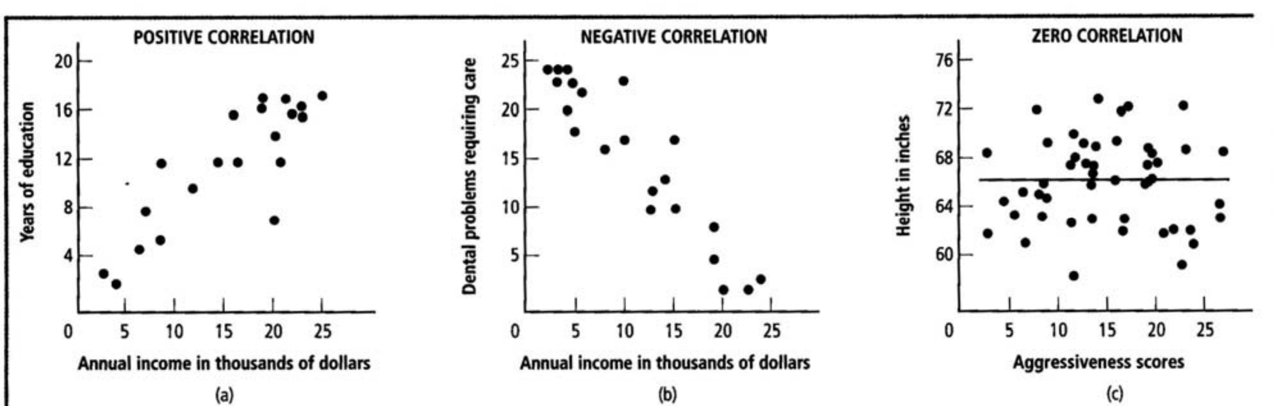


Figure 2.2
CORRELATIONS
Graph (a) shows a positive correlation: In general, income rises with education. Graph (b) shows a negative correlation: In general, the higher people's incomes, the fewer dental problems they have. Graph (c) shows a zero correlation between height and aggressiveness.

A negative correlation means that high values of one variable are associated with *low* values of the other. Figure 2.2b shows a negative correlation between average income and the incidence of dental disease for groups of 100 families. Each dot represents one group. In general, as you can see, the higher the income, the fewer the dental problems. In the automobile business, the older the car, the lower the price (except for antiques and models favored by collectors). As for human beings, in general, the older adults are, the fewer miles they can run, the fewer crimes they are likely to commit, and the fewer hairs they have on their heads. And remember that correlation between hours spent watching TV and grade-point averages? It's a negative one: Lots of hours in front of the television are associated with lower grades (Potter, 1987; Ridley-Johnson, Cooper, & Chance, 1983). See whether you can think of other variables that are negatively correlated. Remember that a negative correlation means that a relationship exists: the *more* of one thing, the *less* of another. If there is no relationship between two variables, we say that they are *uncorrelated* (see Figure 2.2c). Shoe size and IQ scores are uncorrelated.

The statistic used to express a correlation is called the coefficient of correlation. This number conveys both the size of the correlation and its direction. A perfect positive correlation has a coefficient of +1.00, and a perfect negative correlation has a coefficient of -1.00. Suppose you weighed ten people and listed them from lightest to heaviest, then measured their heights and listed them from shortest to tallest. If the names on the two lists were in exactly the same order, the correlation between weight and height would be +1.00. If the correlation between two variables is +.80, it means that they are strongly related. If the correlation is -.80, the relationship is just as strong, but it is negative. When there is no association between two variables, the coefficient is zero or close to zero.

Cautions about Correlations

Correlational findings are common in psychology and are often reported in the news. But beware; many supposed "correlations" in the media or on the Internet are based on rumor and anecdote, and turn out to be small, nonexistent, or meaningless. For example, when some parents of autistic children reported that their children first showed symptoms of the disorder after being vaccinated for measles, mumps, and rubella, many parents panicked and refused to have their children immunized. Subsequent studies, however, showed that the supposed "correlation" between vaccinations and autism had no scientific justification; it was just coincidence. Children are vaccinated at about the time that autistic symptoms become apparent (Ashrof, 2001; Vastag, 2001).

Keep in mind, too, that correlations, even when they are real, can be misleading because *a correlation does not show causation.* It is easy to assume that if variable A predicts variable B, A must be causing B--that is, making B happen--but that is not necessarily so. The number of storks nesting in some European villages is positively correlated with the number of human births in those villages. Therefore, knowing when the storks nest allows you to predict when more births than usual will occur. But that doesn't mean that storks bring babies or that babies attract storks! Human births seem to he somewhat more frequent at certain times of the year (you might want to speculate on the reasons), and the peaks just happen to coincide with the storks' nesting periods.

No one would assume, of course, that storks and babies "cause" each other. But in other cases, unwarranted conclusions about causation are more tempting. For example, children's television watching is moderately correlated with their aggressiveness. What does this positive correlation mean? One answer is that violent programs make some children more aggressive; indeed, many psychologists have come to that conclusion (Bushman & Anderson, 2001). But another answer is that aggressive children are drawn to television violence. And a third answer is that some other factor, such as growing up in a family where physical violence is common, accounts for *both* children's aggressiveness and their attraction to violent programs. Indeed, the association between TV violence and aggression is not a simple one.

Similarly, the negative correlation between TV watching and grades might exist because heavy TV watchers have less time to study, but it is also possible that they have some personality trait that causes an attraction to TV and an aversion to studying, or that they use TV as an escape when their grades are low, or that TV is especially appealing to people who are not academically inclined . . . you get the idea.

The moral of the story: When two variables are associated, one variable may or may not be causing the other.

Experiments: Hunting for Causes

Researchers gain plenty of illuminating information from descriptive studies, but when they want to actually track down the causes of behavior, they rely heavily on the experimental method. An experiment allows the researcher to *control*, or manipulate, the situation being studied. Instead of being a passive recorder of what is going on, the researcher actively does something that he or she believes will affect people's behavior and then observes what happens. These procedures allow the experimenter to draw conclusions about cause and effect--about what causes what.

Experimental Variables

Imagine that you are a psychologist and you come across reports suggesting that cigarette smoking improves reaction time on simple tasks. You have a hunch that nicotine has the opposite effect, however, when the task is as complex and demanding as driving a car. You know that on average, smokers have more car accidents than nonsmokers. But you realize that this relationship does not prove that smoking *causes* accidents. Smokers may simply be greater risk-takers than nonsmokers, whether the risk is lung cancer or trying to beat a red light. Or perhaps the distraction of lighting up accounts for the increased accident risk, rather than smoking itself. So you decide to do an experiment to test your hypothesis.

In a laboratory, you ask smokers to "drive" using a computerized driving simulator equipped with a stick shift and a gas pedal. The object, you tell them, is to maximize the distance covered by driving as fast as possible on a winding road while avoiding rear-end collisions. At your request, some of the subjects smoke a cigarette immediately before climbing into the driver's seat. Others do not. You are interested in comparing how many collisions the two groups have. The basic design of this experiment is illustrated in Figure 2.3, which you may want to refer to as you read the next few pages.

The aspect of an experimental situation manipulated or varied by the researcher is known as the independent variable. The reaction of the subjects--the behavior that the researcher tries to predict--is the dependent variable. Every experiment has at least one independent and one dependent variable. In our example, the independent variable is nicotine use: one cigarette versus none. The dependent variable is the number of collisions.

Ideally, everything in the experimental situation *except* the independent variable is held constant--that is, kept the same for all participants. You would not have some people use a stick shift and others an automatic, unless shift type were an independent variable. Similarly, you would not have some people go through the experiment alone and others perform in front of an audience. Holding everything but the independent variable constant ensures that whatever happens is due to the researcher's manipulation and nothing else. It allows you to rule out other interpretations.

Understandably, students often have trouble keeping independent and dependent variables straight. You might think of it this way: The dependent variable-- the outcome of the *study--depends* on the independent variable. When psychologists set up an experiment, they think, "If I do X, the subjects in my study will do Y." The "X" represents manipulation of the independent variable; the "Y" represents the dependent variable: Most variables may be either independent or dependent, depending on what the experimenter wishes to find out. If you want to know whether eating chocolate makes people nervous, then the amount of chocolate eaten is the independent variable. If you want to know whether feeling nervous makes people eat chocolate, then the amount of chocolate eaten is the dependent variable.

Most variables may be either independent or dependent, depending on what the experimenter wishes to find out. If you want to know whether eating chocolate makes people nervous, then the amount of chocolate eaten is the independent variable. If you want to know whether feeling nervous makes people eat chocolate, then the amount of chocolate eaten is the dependent variable.

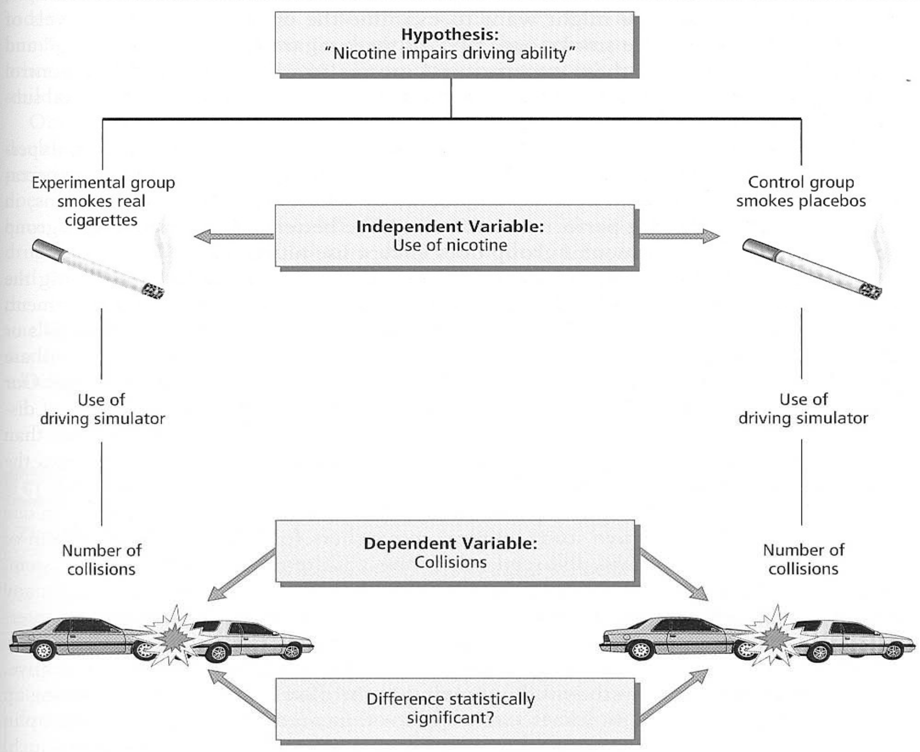


Figure 2.3
DO SMOKING AND DRIVING MIX?
The text describes this experimental design to test the hypothesis that nicotine in cigarettes impairs driving skills.

Experimental and Control Conditions

Experiments usually require both an experimental condition and a comparison, or control condition. In the control condition, subjects are treated exactly as they are in the experimental condition, except that they are not exposed to the same treatment, or manipulation of the independent variable. Without a control condition, you cannot be sure that the behavior you are interested in would not have occurred anyway, even without your manipulation. In some studies, the same subjects can be used in both the control and the experimental conditions; they are said to serve as their own controls. In other studies, subjects are assigned to either an *experimental group* or a *control group.*

In our nicotine experiment, the people who smoke before driving make up the experimental group, and those who refrain from smoking make up the control group. We want these two groups to be roughly the same in terms of average driving skill. It would not do to start out with a bunch of reckless roadrunners in the experimental group and a bunch of tired tortoises in the control group. We probably also want the two groups to be similar in age, education, smoking history, and other characteristics so that none of these variables will affect our results. One way to accomplish this is to use random assignment of people to one group or another--for example, by randomly assigning them numbers and putting those with even numbers in one group and those with odd numbers in another. If we have enough participants in our study, individual characteristics that could possibly affect the results are likely to be roughly balanced in the two groups, so we can ignore them.

Sometimes, researchers use several experimental or control groups. For example, in our nicotine study, we might want to examine the effects of different levels of nicotine by having people smoke one, two, or three cigarettes before "driving," and then comparing each of these experimental groups to each other and to a control group of nonsmokers as well. For now, however, let's focus just on experimental subjects who smoked one cigarette.

We now have two groups. We also have a problem. In order to smoke, the experimental subjects must light up and inhale. These acts might set off certain expectations--of feeling relaxed, nervous, confident, or whatever. These expectations, in turn, might affect driving performance. It would be better to have the control group do everything the experimental group does except use nicotine.

Therefore, we will change the experimental design a bit. Instead of having the control subjects refrain from smoking, we will give them a placebo, a fake treatment. Placebos, which are critical when testing new drugs, often take the form of pills or injections containing no active ingredients. In our study, we will use phony cigarettes that taste and smell like the real thing but contain no nicotine. Our control subjects will not know their cigarettes are fake and will have no way of distinguishing them from real ones. Now if they have substantially fewer collisions than the experimental group, we will feel safe in concluding that nicotine increases the probability of an auto accident.

Control groups, by the way, are also important in nonexperimental studies. In one highly publicized descriptive study, clinical researchers followed the development of 80 children whose parents divorced when the children were. young (Wallerstein, Lewis, & Blakeslee, 2000). The children, who are now in their thirties, revealed many psychological problems and difficulties in relationships, and the researchers interpreted these results as a strong indictment of divorce. But these children all grew up in one extremely affluent county in California, so the sample was not representative. Even more important, the study did not include control subjects: children growing up with unhappy parents in an intact family or, for that matter, children growing up in happy intact families. Without such control groups, the results cannot tell us much. Perhaps growing up with parents who are constantly fighting and abusing one another is worse than having one's parents divorce. For the record, better designed, large-scale studies find that the psychological effects of divorce on children depend on many factors, such as the amount of conflict between the parents before and after the separation (Amato, 1994; Amato & Keith, 1991). Most children of divorce overcome their problems and go on to live well-adjusted lives.

Experimenter Effects

Because expectations can influence the results of a study, participants should not know whether they are in an experimental or a control group. When this is so, the experiment is said to be a single-blind study. But subjects are not the only ones who bring expectations to the laboratory; so do researchers. And researchers' expectations and hopes for a particular result may cause them to inadvertently influence the participants' responses through facial expressions, posture, tone of voice, or some other cue.

Many years ago, Robert Rosenthal (1966) demonstrated how powerful such experimenter effects can be. He had students teach rats to run a maze. Half the students were told that their rats had been bred to be "maze bright," and half were told that their rats had been bred to be "maze dull." In reality, there were no genetic differences between the two groups of rats, yet the supposedly brainy rats actually did learn the maze more quickly, apparently because of the way the students treated them. If an experimenter's expectations can affect a rodent's behavior, reasoned Rosenthal, surely they can affect a human being's. He went on to demonstrate this point in many other studies (Rosenthal, 1994). Even an experimenter's friendly smile (or lack of one) can affect people's responses' in a study.

One solution to the problem of experimenter effects is to do a double-blind study. In such a study, the person running the experiment, the one having actual contact with the subjects, also does not know which subjects are in which groups until the data have been gathered. Double-blind procedures are standard in drug research. Different doses of a drug are coded in some way, and the person administering the drug is kept in the dark about the code's meaning until after the experiment.

Advantages and Limitations of Experiments

Because experiments allow conclusions about cause and effect, and because they permit researchers to distinguish real effects from placebo effects, they have long been the method of choice in psychology. However, like all methods, the experiment has its limitations. *Just* as in other kinds of studies, the participants are not always representative of the larger population. Most volunteers in academic experiments are college students, who differ in many ways from people who are not in school. Moreover, in an experiment, the researcher determines which questions are asked and which behaviors are recorded, and the participants try to do as they are told. In their desire to cooperate with the experimenter or present themselves in a positive light, they may act in ways that they ordinarily would not (Kihlstrom, 1995).

Thus, research psychologists confront a dilemma: The more control they exercise over the situation, the more artificial the situation--and the results obtained from it--may be. For this reason, many psychologists have called for more field research, the careful study of behavior in natural contexts such as schools and the workplace, using both descriptive and experimental methods. For example, a psychologist interested in ways of reducing prejudice might study that problem not just in the laboratory, but also in offices and schools. Field research on prejudice has led to successful strategies for reducing hostility among children and adults of different ethnicities.

Every research method has both its strengths and its weaknesses.

Evaluating the Findings

If you are a psychologist who has just done an observational study, a survey, or an experiment, your work has just begun. Once you have some results in hand, you must do three things with them: (1) describe them, (2) assess how reliable and meaningful they are, and (3) figure out how to explain them.

Descriptive Statistics: Finding Out What's So

Let's say that 30 people in the nicotine experiment smoked real cigarettes, and 30 smoked placebos. We have recorded the number of collisions for each person on the driving simulator. Now we have 60 numbers. What can we do with them?

The first step is to summarize the data. The world does not want to hear how many collisions each person had. It wants to know what happened in the nicotine group as a whole, compared to what happened in the control group. To provide this information, we need numbers that sum up our data. Such numbers, known as descriptive statistics, are often depicted in graphs and charts.

A good way to summarize the data is to compute group averages. The most commonly used type of average is the arithmetic mean. The mean is calculated by adding up all the individual scores and dividing the result by the number of scores. We can compute a mean for the nicotine group by adding up the 30 collision scores and dividing the sum by 30. Then we can do the same for the control group. Now our 60 numbers have been boiled down to 2. For the sake of our example, let's assume that the nicotine group had an average of 10 collisions, whereas the control group's average was only 7.

At this point in our nicotine study, we have one group with an average of 10 collisions and another with an average of 7. Should we break out the champagne? Try to get on TV? Call our mothers? Better hold off. Perhaps if one group had an average of 15 collisions and the other an average of 1, we could get excited. But rarely does a psychological study hit you between the eyes with a sensationally clear result. In most cases, there is some possibility that the difference between the two groups was due simply to chance. Despite all our precautions, perhaps the people in the nicotine group just happened to be a little more accident-prone, and their behavior had nothing to do with the nicotine.

Inferential Statistics: Asking “So What?”

To find out how impressive the data are, psychologists use inferential statistics. These statistics do not merely describe or summarize the data; they permit a researcher to draw *inferences* (conclusions based on evidence) about how meaningful the findings are. Like descriptive statistics, inferential statistics involve the application of mathematical formulas to the data.

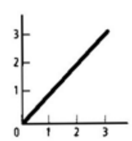
The most commonly used inferential statistics are significance tests, which tell researchers how likely a result was to have occurred by chance. In our nicotine study, a significance test will tell us how likely it is that the difference between the nicotine group and the placebo group occurred by chance. It is not possible to rule out chance entirely, but if the likelihood that a result occurred by chance is extremely low, we can say that the result is *statistically significant.* This means that the probability that the difference is "real" is overwhelming -- not certain, mind you, but overwhelming.

By convention, psychologists consider a result to be significant if it would be expected to occur by chance 5 or fewer times in 100 repetitions of the study. Another way of saying this is that the result is significant at the .05 ("point oh five") level. If the difference could be expected to occur by chance in 6 out of 100 studies, we would have to say that the results failed to support the hypothesis that the difference we obtained might well have occurred merely by chance--although we might still want to do further research to be sure. You can see that psychologists refuse to be impressed by just any old result.

Statistically significant results allow psychologists to make general predictions about human behavior. These predictions are usually stated as probabilities ("On average, we can expect 60 percent of all students to do X, Y, or Z"). However, they usually do not tell us with any certainty what a particular individual will do in a particular situation. Probabilistic results are typical not only in psychology but in all of the sciences. Medical research, for example, can tell us that the odds are high that someone who smokes will get lung cancer, but because many variables interact to produce any particular case of cancer, research cannot tell us for sure whether Aunt Bessie, a two-pack-a-day smoker, will come down with the disease.

By the way, a nicotine study similar to our hypothetical example was actually done some years ago, using somewhat more complicated procedures (Spilich, June, & Renner, 1992). Smokers who lit up before driving got a little farther on the simulated road, but they also had significantly more rear-end collisions on average (10.7) than did temporarily abstaining smokers (5.2) or nonsmokers (3.1). The lead researcher in this study told us that after hearing about these findings, the head of Federal Express banned smoking on the job among all of the company's drivers.

Figure 1. Advantages and Disadvantages of Various Research Methods

Method	Advantages	Disadvantages
Case study 	Good source of hypotheses. Provides in-depth information on individuals. Unusual cases can shed light on situations or problems that are unethical or impractical to study in other ways.	Vital information may be missing, making the case hard to interpret. The person's memories may be selective or inaccurate. The individual may not be representative or typical.
Naturalistic observation 	Allows description of behavior as it occurs in the natural environment. Often useful in first stages of a research program.	Allows researcher little or no control of the situation. Observations may be biased. Does not allow firm conclusions about cause and effect.
Survey 	Provides a large amount of information on large numbers of people.	If sample is nonrepresentative or biased, it may be impossible to generalize from the results. Responses may be inaccurate or untrue.
Correlational study 	Shows whether two or more variables are related. Allows general predictions.	Does not permit identification of cause and effect.
Experiment 	Allows researcher to control the situation. Permits researcher to identify cause and effect, and to distinguish placebo effects from treatment effects.	Situation is artificial, and results may not generalize well to the real world. Sometimes difficult to avoid experimenter effects.

From: <u>Psychology</u> by Wade and Tavris, 7th Ed. 2005. Pearson Publishing. pp 40 – 57.