LEARNING OBJECTIVES

1. List the three criteria for establishing a causal relationship and the two cautions that can improve understanding of a causal connection.
2. Contrast the strengths and weaknesses of dealing with nonspuriousness through statistical control and through randomization.
3. Explain the meaning of the expression “correlation does not prove causation.”
4. Name two challenges to using experimental designs and two difficulties with identifying idiographic causal explanations.
5. Name and illustrate the three different quasi-experimental designs.
6. Define the individual and group units of analysis and explain the role they play in the ecological and reductionist fallacies.

What Do We Mean by Causation?

At the end of 2013, almost 2.5 million people in the United States were incarcerated in prisons or local jails (Glaze & Kaeble, 2014). The large prison population coupled with lawsuits has prompted most correctional institutions to begin classifying inmates into different security levels (e.g., minimum and maximum) based on objective criteria such as the severity of their offenses, previous records, and so forth. Obviously, the security level of an institution in which an inmate is classified will affect his or her incarceration experience. For example, someone who is assigned to a maximum security prison instead of one with a lower level of security will also have differential access to things such as...
Causal Explanation

A cause is an explanation for some characteristic, attitude, or behavior of groups, individuals, or other entities (such as families, organizations, or cities) or for events. Most social scientists seek causal explanations that reflect tests of the types of hypotheses with which you are familiar (recall Chapter 2). In these tests, the independent variable is the presumed cause and the dependent variable is the potential effect. For example, does problem-oriented policing (independent variable) reduce violent crime (dependent variable)? Does experiencing abuse as a child (independent variable) increase the likelihood that the person will be a violent adult (dependent variable)? This type of causal explanation is termed nomothetic.

A different type of cause is the focus of some qualitative research and our everyday conversations about causes. In this type of causal explanation, termed idiographic, individual events or the behavior of individuals are explained with a series of related, prior events. For example, you might explain a particular crime as resulting from several incidents in the life of the perpetrator that resulted in a tendency toward violence, coupled with stress resulting from a failed marriage and a chance meeting with a suitable target.

Quantitative (Nomothetic) Causal Explanation

A nomothetic causal explanation is one involving the belief that variation in an independent variable will be followed by variation in the dependent variable, when all other things are equal (ceteris paribus) or when all other potentially influential conditions and factors are taken into consideration. For instance, researchers might claim that the likelihood of committing violent crimes is higher for individuals who were abused as children than it would be if these same individuals had not been abused as children. Or, researchers might claim that the likelihood of committing violent crimes is higher for individuals exposed to media violence than it would be if these same individuals had not been exposed to media violence. The situation as it would have been in the absence of variation in the independent variable is termed the counterfactual (see Exhibit 6.1).

Of course, the fundamental difficulty with this perspective is that we never really know what would have happened at the same time to the same people (or groups, cities, etc.) if the independent variable had not varied because it did. We cannot run real-life scenarios (King, Keohane, & Verba, 1994). We could observe the aggressiveness of people's behavior before and after they were exposed to media violence. But this comparison involves an earlier time period, when by definition, the people and their circumstances were not exactly the same.

Fortunately, we can design research to create conditions that are comparable indeed so that we can confidently assert our conclusions ceteris paribus. We can examine the impact on the dependent variable of variation in the independent variable alone, even though we will not be able to compare the same people at the same time in exactly the same circumstances except for the variation mental health services, drug treatment programs, and vocational training (Brennan & Austin, 1997). But is the classification of inmates also related to their behavior while incarcerated? Do those assigned to maximum security prisons engage in more misconduct compared to inmates assigned to less secure facilities? How could you answer this question? If you compared rates of misconduct across prison settings, you would not have the answer because the inmates may have been very different at their time of incarceration. As such, any differences you observe in misconduct could be attributable to these “before incarceration” differences, not to the type of facility in which they are housed. As you can see, establishing causation is more difficult than it appears.

Nomothetic causal explanation  A type of causal explanation involving the belief that variation in an independent variable will be followed by variation in the dependent variable, when all other things are equal.

Causal effect (nomothetic perspective)  When variation in one phenomenon, an independent variable, leads to or results, on average, in variation in another phenomenon, the dependent variable.

Example of a nomothetic causal effect  Individuals arrested for domestic assault tend to commit fewer subsequent assaults than do similar individuals who are accused in the same circumstances but not arrested.

Ceteris paribus  Latin term meaning “all other things being equal.”

Counterfactual  The outcome that would have occurred if the subjects who were exposed to the treatment actually were not exposed but otherwise had had identical experiences to those they underwent during the experiment.
in the independent variable. And by knowing the ideal standard of comparability, we can improve our research designs and strengthen our causal conclusions even when we cannot come so close to living up to the meaning of *ceteris paribus*.

Quantitative researchers seek to test nomothetic causal explanations with either experimental or nonexperimental research designs. However, the way in which experimental and nonexperimental designs attempt to identify causes differs quite a bit. It is very hard to meet some of the criteria for achieving valid nomothetic causal explanations using a nonexperimental design. Most of the rest of this chapter is devoted to a review of these causal criteria and a discussion of how experimental and nonexperimental designs can help establish them.

### Qualitative (Idiographic) Causal Explanation

The other meaning of the term *cause* is one that we have in mind very often in everyday speech. This is *idiographic causal explanation*: the concrete, individual sequence of events, thoughts, or actions that resulted in a particular outcome for a particular individual or that led to a particular event (Hage & Meeker, 1988). An idiographic explanation also may be termed an *individualist* or a *historicist explanation*.

A *causal effect* from an *idiographic perspective* includes statements of initial conditions and then relates a series of events at different times that led to the outcome, or causal effect. This narrative, or story, is the critical element in an idiographic explanation, which may therefore be classified as narrative reasoning (Richardson, 1995). Idiographic explanations focus on particular social actors, in particular social places, at particular social times (Abbott, 1992). Idiographic...
explanations are also typically very concerned with context, with understanding the particular outcome as part of a larger set of interrelated circumstances. Idiographic explanations thus can be termed holistic.

Elijah Anderson’s (1999) field research in a poor urban community produced a narrative account of how drug addiction can result in a downward slide into residential instability and crime:

When addicts deplete their resources, they may go to those closest to them, drawing them into their schemes. . . . The family may put up with the person for a while. They provide money if they can. . . . They come to realize that the person is on drugs. . . . Slowly the reality sets in more and more completely, and the family becomes drained of both financial and emotional resources. . . . Close relatives lose faith and begin to see the person as untrustworthy and weak. Eventually the addict begins to “mess up” in a variety of ways, taking furniture from the house [and] anything of value. . . . Relatives and friends begin to see the person . . . as “out there” in the streets. . . . One deviant act leads to another. (pp. 86–87)

An idiographic explanation like Anderson’s pays close attention to time order and causal mechanisms. Nonetheless, it is difficult to make a convincing case that one particular causal narrative should be chosen over an alternative narrative (Abbott, 1992). Does low self-esteem result in vulnerability to the appeals of drug dealers, or does a chance drug encounter precipitate a slide in self-esteem? The prudent causal analyst remains open to alternative explanations.

Idiographic explanation is deterministic, focusing on what caused a particular event to occur or what caused a particular case to change. As in nomothetic explanations, idiographic causal explanations can involve counterfactuals by trying to identify what would have happened if a different circumstance had occurred. But unlike nomothetic explanations, in idiographic explanations, the notion of a probabilistic relationship, an average effect, does not really apply. A deterministic cause has an effect only in the case under consideration. We focus on methods for examining idiographic causation more closely in Chapter 8.

Criteria for Nomothetic Causal Explanations

Mark Twitchell wanted to be a filmmaker and become famous. One of the short movies he made was about a serial killer. Twitchell also was a big fan of the TV show Dexter, a drama about a serial killer. In 2008, he advanced this fiction to real life when he posed as a woman on a dating Web site to lure Johnny Altinger on a date. When Altinger showed up for the date on October 8, 2008, he was killed and dismembered. Fortunately, the murder was discovered before Twitchell could kill again. Not surprisingly, after his arrest, Twitchell became known as the “Dexter Killer” (Edmonton Journal, 2015). As frequently happens, some attributed Twitchell’s violence to media portrayals of violence, in this case, to the series Dexter. How would you evaluate this claim? What evidence do we need to develop a valid conclusion about a hypothesized causal effect? Imagine a friend saying, after reading about the incident, “See, media violence causes people to commit crimes.” Of course, after reading Chapter 1 you would not be so quick to jump to such a conclusion. “Don’t overgeneralize,” you would remind yourself. When your friend insists, “But I recall that type of thing happening before,” you might even suspect selective observation. As a blossoming criminological researcher, you now know that if we want to have confidence in the validity of our causal statements, we must meet a higher standard.

How research is designed influences our ability to draw causal conclusions. In this section, we will introduce the features that need to be considered in a research design to evaluate how well it can support nomothetic causal conclusions.

Five criteria must be considered when deciding whether a causal connection exists. When a research design leaves one or more of the criteria unmet, we may have some important doubts about causal assertions the researcher may have made. The first three of the criteria are generally considered the necessary and most important basis for identifying a nomothetic causal effect: empirical association, appropriate

<table>
<thead>
<tr>
<th>Context</th>
<th>A focus of causal explanation; a particular outcome is understood as part of a larger set of interrelated circumstances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example of an idiographic causal effect</td>
<td>An individual is neglected by his parents. He comes to distrust others, has trouble maintaining friendships, has trouble in school, and eventually gets addicted to heroin. To support his habit, he starts selling drugs and is ultimately arrested and convicted for drug trafficking</td>
</tr>
</tbody>
</table>
time order, and nonspuriousness. The other two criteria, identifying a causal mechanism and specifying the context in which the effect occurs, can also considerably strengthen causal explanations, although many do not consider them as requirements for establishing a causal relationship.

**Conditions necessary for determining causality:**

1. Empirical association
2. Appropriate time order
3. Nonspuriousness

**Conditions important in specifying causal relationships:**

1. Mechanism
2. Context

---

**Case Study**

**Media Violence and Violent Behavior**

We will use Brad Bushman’s (see Bushman & Huesmann, 2012, for review) experiments on media violence and aggression to illustrate the five criteria for establishing causal relationships. Bushman’s study focused in part on this specific research question: Do individuals who view a violent videotape act more aggressively than individuals who view a nonviolent videotape?

Undergraduate psychology students were recruited to watch a 15-minute videotape in a screening room, one student at a time. Half of the students watched a movie excerpt that was violent, and half watched a nonviolent movie excerpt. After viewing the videotape, the students were told that they were to compete with another student, in a different room, on a reaction-time task. When the students saw a light cue, they were to react by trying to click a computer mouse faster than their opponent. On a computer screen, the students set a level of radio static that their opponents would hear when the opponents reacted more slowly. The students themselves heard this same type of noise when they reacted more slowly than their opponents at the intensity level supposedly set by their opponents.

Each student in the study participated in 25 trials, or competitions, with the unseen opponent. Their aggressiveness was operationalized as the intensity of noise that they set for their opponents over the course of the 25 trials. The louder the noise level they set, the more aggressively they were considered to be behaving toward their opponents. The question that we will focus on first is whether students who watched the violent video behaved more aggressively than those who watched the nonviolent video.

**Association**

Exhibit 6.2 displays the association that Brad Bushman found between watching a violent videotape and aggressive behavior (Bushman & Huesmann, 2012). Students who watched a violent videotape in his lab administered more intense noise to an opponent than those who watched a nonviolent videotape. Thus, variation in exposure to media violence is associated with a likelihood of exhibiting aggressive behavior. A change in the independent variable is associated with—correlated with—a change in the dependent variable. By contrast, if there is no association between two variables, there cannot be a causal relationship.
Time Order

Association is a necessary criterion for establishing a causal effect, but it is not sufficient on its own. We must also ensure that the variation in the independent variable came before variation in the dependent variable—the cause must come before the presumed effect. This is the criterion of time order. Bushman's original experiment (1995) satisfied this criterion because he controlled the variation in the independent variable: All the students saw the movie excerpts (which varied in violent content) before their level of aggressiveness was measured. As you can imagine, we cannot be so sure about time order when we use a survey or some other observation done at one point in time. For example, if we find that neighborhoods with higher levels of disorder have higher crime rates, we can't be sure that the level of disorder came first and that it led to more crime. Maybe higher crime rates made residents too fearful to keep things in order. Without longitudinal data or other clues to time order, we just don't know.

Nonspuriousness

Even when research establishes that two variables are associated and that variation in the independent variable precedes variation in the dependent variable, we cannot be sure we identified a causal relationship between the two variables. Have you
heard the old adage among researchers, “Correlation does not prove causation”? It is meant to remind us that an association between two variables might be caused by something else. If we measure children’s shoe sizes and their academic knowledge, for example, we will find a positive association. However, the association results from the fact that older children tend to have larger feet as well as more academic knowledge. As it turns out, shoe size does not cause increased knowledge or vice versa. We would say that the association between shoe size and knowledge is spurious (meaning apparently but not actually valid—that is, false).

Before we conclude that variation in an independent variable causes variation in a dependent variable, we must have reason to believe that the relationship is nonspurious. Nonspuriousness is a relationship between two variables that is not due to variation in a third variable (see Exhibit 6.3).

Does Bushman et al.’s (2012) claim of a causal effect rest on any stronger ground? To evaluate nonspuriousness, you need to know about one more feature of his experiment. He assigned students to watch either the violent video or the nonviolent video randomly—that is, by the toss of a coin. Because he used random assignment, the characteristics and attitudes that students already possessed when they were recruited for the experiment could not influence which video they watched. As a result, the students’ characteristics and attitudes could not explain why one group reacted differently from the other after watching the videos. In fact, because Bushman used 296 students in his experiment, it is highly unlikely that the violent video group and the nonviolent video group differed in any relevant way at the outset, even on the basis of chance. This experimental research design meets the criterion of nonspuriousness.

Bushman’s conclusion that viewing video violence causes aggressive behavior thus rests on firm ground indeed.

Mechanism

Confidence in a conclusion that two variables have a causal connection will be strengthened if a mechanism—some discernible means of creating a connection—can be identified (Cook & Campbell, 1979, p. 35; Marini & Singer, 1988). Many social scientists believe that a causal explanation is not adequate until a causal mechanism is identified—what process or mechanism actually is responsible for the relationship between the independent and dependent variables.

Exhibit 6.3 A Spurious Relationship

Spurious relationship

View the series
*Dexter*

Commit violent crime

The extraneous variable creates the spurious relationship

Feel enraged against society

Watch violent movie

Commit violent crime

Copyright ©2018 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Bushman and Huesmann (2012) did not empirically identify a causal mechanism in their experiment, though Bushman did suggest a possible causal mechanism for the effect of watching violent videos. Before we can explain this causal mechanism, we have to tell you about one more aspect of Bushman's research. He was not interested simply in whether viewing violent films resulted in aggressive behavior. Actually, his primary hypothesis was that individuals who were predisposed to aggression before the study began would be more influenced by a violent film than individuals who were not aggressive at the outset. And that is what happened: Individuals who were predisposed to aggression became more aggressive after watching Bushman's violent video, but individuals who were not predisposed to aggression did not become more aggressive.

After the experiment, Bushman and Huesmann (2012) proposed a causal mechanism to explain why aggressive individuals became even more aggressive after watching the film:

High trait aggressive individuals [people predisposed to aggression] are more susceptible to the effects of violent media than are low trait aggressive individuals because they possess a relatively large network of aggressive associations that can be activated by violent cues. Habitual exposure to television violence might be partially responsible. (p. 959)

Note that this explanation relies more on speculation than on the actual empirical evidence from this particular experiment. Nonetheless, by proposing a reasonable causal mechanism that connects the variation in the independent and dependent variables, this strengthened the argument for the causal validity of their conclusions.

**Context**

In the social world, it is virtually impossible to claim that one and only one independent variable is responsible for causing or affecting a dependent variable. Stated another way, no cause can be separated from the larger context in which it occurs. A cause is really only one of a set of interrelated factors required for the effect (Hage & Meeker, 1988; Papineau, 1978). When relationships among variables differ across geographic units such as counties or across other social settings, or even between different types of individuals, researchers say there is a contextual effect. Identification of the context in which a causal relationship occurs can help us to understand that relationship.

In a classic study of children's aggressive behavior in response to media violence, Albert Bandura, Dorothea Ross, and Sheila Ross (1963) examined several contextual factors. For example, they found that children reacted more aggressively after observing men committing violent acts than after observing women committing these same acts. Bandura and colleagues strengthened their conclusions by focusing on a few likely contextual factors. Specifying the context for a causal effect helps us understand that effect, but it is a process that can never really be complete. We can always ask what else might be important. In which country was the study conducted? What are the ages of the study participants? We need to carefully review the results of prior research and the implications of relevant theory to determine what contextual factors are likely to be important in a causal relationship.

### Why Experiment?

How research is designed influences our ability to draw causal conclusions. Obviously, if you conclude that playing violent video games causes violent behavior after watching your 8-year-old nephew playing a violent video game and then hitting his 4-year-old brother, you would be on shaky empirical ground. In this section, we will introduce features that need to be considered in a research design to evaluate how well it can support nomothetic causal conclusions.

### True Experiments

Experimental research provides the most powerful design for testing causal hypotheses because it allows us to confidently establish the first three criteria for causality—association, time order, and nonspuriousness. True experiments have at least three features that help us meet these criteria:
1. Two comparison groups—one receiving the experimental condition (e.g., treatment or intervention), termed the experimental group, and the other receiving no treatment/intervention or another form thereof, termed the control group.

2. Random assignment to the two (or more) comparison groups.

3. Assessment of change in the dependent variable for both groups after the experimental condition has been applied. This is usually called a posttest. We can determine whether an association exists between the independent and dependent variables in a true experiment because two or more groups differ in terms of their value on the independent variable. One group, the experimental group, receives some “treatment” that is a manipulation of the value of the independent variable. In a simple experiment, there may be one other group that does not receive the treatment; it is termed the control or comparison group.

Let’s consider the Bushman experiment in detail (see the simple diagram in Exhibit 6.4). Does watching a violent video lead to aggressive behavior? Imagine a simple experiment. Suppose you believe that watching violent movies leads people to be aggressive, commit crimes, and so on. But other people think that violent media has no effect and that it is people who are already predisposed to violence who seek out violent movies to watch. To test your research hypothesis (“Watching violent movies causes aggressive behavior”), you need to compare two randomly assigned groups of subjects, a control group and an experimental group.

First, it is crucial that the two groups be more or less equal at the beginning of the study. If you let students choose which group to be in, the more violent students may pick the violent movie, hoping, either consciously or unconsciously, to have their aggressive habits reinforced. If so, your two groups won’t be equivalent at the beginning of the study. As such, any difference in their aggressiveness may be the result of that initial difference (a source of spuriousness), not whether they watched the violent video. You must randomly sort the students into the two different groups. You can do this by flipping a coin for each one of them, pulling names out of a hat, or using a random number table as described in the previous chapter. In any case, the subjects themselves should not be free to choose nor should you (the experimenter) be free to put them into whatever group you want.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Placement in group</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate students</td>
<td>Random assignment</td>
<td>Experimental group</td>
<td>Degree of aggression</td>
<td>Watched violent video</td>
<td>Degree of aggression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control group</td>
<td>Degree of aggression</td>
<td>Watched nonviolent video</td>
<td>Degree of aggression</td>
</tr>
</tbody>
</table>

**Exhibit 6.4 Experimental Design Used in the Bushman Research**

**True experiment** Experiment in which subjects are assigned randomly to an experimental group that receives a treatment or other manipulation of the independent variable and a comparison group that does not receive the treatment or receives some other manipulation; outcomes are measured in a posttest.

**Experimental group** In an experiment, the group of subjects that receives the treatment or experimental manipulation.

**Control or comparison group** The group of subjects who are either exposed to a different treatment than the experimental group or who receive no treatment at all.
Note that the random assignment of subjects to experimental and comparison groups is not the same as random sampling of individuals from some larger population (see Exhibit 6.5). In fact, random assignment (randomization) does not help at all to ensure that the research subjects are representative of some larger population; instead, representativeness is the goal of random sampling. What random assignment does—create two (or more) equivalent groups—is useful for ensuring internal validity, not generalizability.

Next, people in the two groups will interact among themselves. Then, the control group will watch a video about gardening while the experimental group will watch a video featuring a lot of violence. Next, both groups will sit and interact again among themselves. At the end, the interactions within both groups before and after the videos will be coded and you will see whether either group increased in aggressiveness. Thus, you may establish association.

**Exhibit 6.5 Random Sampling Versus Random Assignment**

Random sampling (a tool for ensuring generalizability):
Individuals are randomly selected from a population to participate in a study.

Random assignment, or randomization (a tool for ensuring internal validity):
Individuals who are to participate in a study are randomly divided into an experimental group and a comparison group.
Matching is another procedure sometimes used to equate experimental and comparison groups, but by itself, it is a poor substitute for randomization. Matching of individuals in a treatment group with those in a comparison group might involve pairing persons on the basis of similarity of gender, age, year in school, or some other characteristic. The basic problem is that, as a practical matter, individuals can be matched on only a few characteristics. Unmatched and unknown differences between the experimental and comparison groups may still influence outcomes.

These defining features of true experimental designs give us a great deal of confidence that we can meet the three basic criteria for identifying causes: association, time order, and nonspuriousness. However, we can strengthen our understanding of causal connections and increase the likelihood of drawing causally valid conclusions by also investigating causal mechanism and causal context.

Even after establishing the random assignment of experimental and control groups, you may find an association outside the experimental setting, but it won’t establish time order. Perhaps aggressive people choose to watch violent videos, while nonaggressive people do not. So there would be an association but not the causal relation for which we are looking. By controlling who watches the violent video and when, we establish time order. All true experiments have a posttest—that is, a measurement of the outcome in both groups after the experimental group has received the treatment. Many true experiments also have pretests that measure the dependent variable before the experimental intervention. A pretest is exactly the same as a posttest, just administered at a different time. Strictly speaking, though, a true experiment does not require a pretest. When researchers use random assignment, the groups’ initial scores on the dependent variable (or observed effect or behavior) and on all other variables are very likely to be similar. Any difference in outcome between the experimental and comparison groups is therefore likely to be due to the intervention (or to other processes occurring during the experiment), and the likelihood of a difference solely on the basis of chance can be calculated.

An Experiment in Action: Prison Classification and Inmate Behavior

There is wide variability in the criteria used to classify prisoners across the United States. Regardless of how these classifications are made, once these labels are assigned, they have the effect that all labels have: They attach various stigmas and expectations to prisoners. Bench and Allen (2003) state,

An offender classified as maximum security instantly obtains an image of one who is hard to handle, disrespectful of authority, prone to fight with other inmates, and at a high risk for escape. In contrast, an offender classified as medium security is generally regarded as more manageable, less of an escape risk, and not requiring as much supervision as a maximum-security offender. (p. 371)

To examine whether prison classification actually affects inmate behavior, Bench and Allen (2003) obtained a random sample of 200 inmates admitted to the Utah State Prison who had been classified as maximum security following their initial assessment based on the following criteria: severity of current crime, expected length of incarceration, criminal violence history, escape history, prior institutional commitment, age, history of institutional adjustment, and substance abuse history.

From this group, inmates were randomly assigned to either an experimental group, in which inmates were reclassified to medium-security status, or a control group, in which inmates retained their maximum-security status. The independent variable, then, was security classification. The dependent variable was the number of disciplinary infractions or sanctions for violation of prison rules received by each group. The severity of infractions was weighted to control for the severity of the violations (e.g., possession of unauthorized food was weighted lower than assaulting another inmate). The primary hypothesis was that the
experimental group, those reclassified as medium security, would have a lower number of disciplinary infractions compared with the control group, the inmates who retained their maximum-security classification. A diagram depicting the experiment is provided in Exhibit 6.6. Results indicated that inmates reclassified to medium security did not receive a lower number of infractions; both groups received about the same number of disciplinary infractions, regardless of security classification.

Field Experiments in Action: Determining the Effect of Incarceration on Employment

As you have seen, social experiments are not always conducted in a laboratory or controlled environment. In fact, many experiments are conducted out in the real world. Whenever studies utilize the conditions of an experimental method in a real-world setting, they are termed field experiments. All of the studies examining the effects of arrest on future intimate partner assaults discussed in Chapter 2 were field experiments.

One innovative field experiment was conducted by Pager (2007) to determine the effects of incarceration on the likelihood of obtaining employment. This is an extremely important research question because the prison population in the United States has vastly increased over the past 30 years. In addition to the laws barring ex-offenders in some states from obtaining employment in certain sectors, reentering offenders also face other obstacles in finding a job, particularly from the stigma attached to having a record. How could we determine the effects of a formal criminal record on the likelihood of getting a job? Well, we could examine employer attitudes about hiring ex-offenders through a survey, but as you now know, this would not help us isolate a causal relationship between having a record and getting a job. We could interview offenders reentering the community to find out their own experiences, but this would tell us only about a few individuals’ experiences. The best way to determine the effects of a criminal record on employment chances would be to conduct a field experiment, which is what Pager did. Her study is diagrammed in Exhibit 6.7.

Pager (2007) designed a field experiment in which pairs of applicants, one who had a criminal record and one who did not, applied for real jobs. Her study used two male teams of applicants, one composed of two African Americans and one composed of two whites. These individuals were actually college students in Milwaukee, Wisconsin, whom Pager refers to as “testers.” The testers were matched on the basis of age, physical appearance, and general style of self-presentation, and all were assigned fictitious résumés that reflected equivalent levels of education (all had a high school education) and equivalent levels of steady work experience. However, one tester within each team was randomly assigned to have a criminal record and the other was not. The fictitious criminal record consisted of a felony drug conviction and 18 months of served prison time. This assignment rotated each week of the study (e.g., one individual played the job applicant with a record one week, and the other did so the next week) as a check against unobserved differences between team members. Same-race testers (one with a criminal record and one without) applied for the same job, one day apart. The African American team applied for a total of 200 jobs, and the white team applied for a total of 150 jobs.
The primary outcome of the study was the proportion of applications that elicited either callbacks from employers or on-the-spot job offers. The testers went through intensive training to become familiar with their assumed profiles and to respond similarly to potential interview questions. As such, the only difference between the two testers on each race team was that one had a criminal record and the other didn’t. Because there was random assignment to these two conditions and the other characteristics of the testers were essentially the same, the differences observed in the percentage of callbacks between team members can be assumed to be related to the criminal record only and not to other factors.

The results of Pager’s (2007) field experiment were stark. White testers with a criminal record were one-half to one-third less likely to receive a callback from employers, and the effect was even more pronounced for African American applicants. Pager concludes, “Mere contact with the criminal justice system in the absence of any transformative or selective effects severely limits subsequent job prospects. The mark of a criminal record indeed represents a powerful barrier to employment” (p. 145). With such a powerful randomly assigned field experiment, the internal (causal) validity of these findings is strong. The implications of these findings in light of the hundreds of thousands of offenders who attempt to reenter society from prison each year are troubling indeed.

What If a True Experiment Isn’t Possible?

Often, testing a hypothesis with a true experimental design is not feasible. A true experiment may be too costly or take too long to carry out; it may not be ethical to randomly assign subjects to the different conditions or it may be too late to do so. Researchers may instead use quasi-experimental designs that retain several components of experimental design but differ in important details.

In a quasi-experimental design, a comparison group is predetermined to be comparable to the treatment group in critical ways, such as being eligible for the same services or being in the same school cohort (Rossi & Freeman, 1989, p. 313). These research designs are only quasi-experimental because subjects are not randomly assigned to the comparison and experimental groups. As a result, we cannot be as confident in the comparability of the groups as in true experimental designs. Nonetheless, in order to term a research design quasi-experimental, we have to be sure that the comparison groups meet specific criteria.
We will discuss here the two major types of quasi-experimental designs as well as one type—ex post facto (after the fact) control group design—that is often mistakenly termed quasi-experimental. (Other types can be found in Cook & Campbell, 1979, and Mohr, 1992.)

- **Nonequivalent control group designs**: These designs have experimental and comparison groups that are designated before the treatment occurs but are not created by random assignment.

- **Before-and-after designs**: This type of design has a pretest and posttest but no comparison group. In other words, the subjects exposed to the treatment served, at an earlier time, as their own control group.

- **Ex post facto control group designs**: These designs use nonrandomized control groups designated after the fact.

These designs are weaker than true experiments in establishing the nonspuriousness of an observed association—that it does not result from the influence of some third, uncontrolled variable. On the other hand, because these quasi-experiments do not require the high degree of control necessary in order to achieve random assignment, they can be conducted using more natural procedures in more natural settings, so we may be able to achieve a more complete understanding of causal context. In identifying the mechanism of a causal effect, though, quasi-experiments are neither better nor worse than experiments.

---

**Nonequivalent Control Group Designs**

In this type of quasi-experimental design, a comparison group is selected to be as comparable as possible to the treatment group. Two selection methods can be used:

1. **Individual matching**: Individual cases in the treatment group are matched with similar individuals in the comparison group. This can sometimes create a comparison group that is very similar to the experimental group.

2. **Aggregate matching**: In most situations when random assignment is not possible, this second method of matching makes more sense: identifying a comparison group that matches the treatment group in the aggregate rather than trying to match individual cases. This means finding a comparison group that has similar distributions on key variables: the same average age, the same percentage female, and so on.

---

**Case Study**

**The Effectiveness of Drug Courts**

Listwan et al.’s (2003) study of the effectiveness of a drug court on recidivism illustrates a quasi-experimental nonequivalent control group design. Their quasi-experimental design is diagrammed in Exhibit 6.8. Reflecting the priority that policy makers place on controlling drug use and drug-related crime, drug courts have become extremely popular in the United States. They emerged as an alternative to correctional prison and jail-based responses to addicted offenders and generally rely on community-based treatment models. The assumption behind the drug court movement is that drug users and their drug-related crimes will increasingly clog the courts and fill our jails and prisons if their addictions are not remedied. Although drug court programs vary tremendously across jurisdictions, they generally integrate alcohol and drug treatment services with justice system case processing. In addition, they are designed to decrease case-processing time, alleviate the demand of drug-related cases on the court, and decrease jail and prison commitments to drug-related offenders, all of which are supposed to decrease the cost of controlling drug offenders.
Listwan and her colleagues (2003) examined whether participants in the Hamilton County Drug Court program in Cincinnati, Ohio, had lower rates of recidivism for both drug-related and other offenses. Exhibit 6.8 illustrates that there were two groups compared: those who participated in the drug court (experimental group) and those who were eligible but did not receive the drug court treatment services or the additional court supervision (control group). Importantly, the offenders were not randomly assigned to these groups. The researchers simply stated, “Members of [the control group] . . . either refused drug treatment or were refused by the drug court team” (p. 396).

Unfortunately, many evaluations of this nature do not have the ability to employ random assignment, thereby diluting the ability to determine the causal relationship between the treatment and the results. The researchers did examine the potential differences between the two groups and determined that they did not significantly differ in terms of age, race, education, or prior arrest for a drug-related offense, but the experimental group had a higher number of women and people with other prior records not related to drugs. Arrest and incarceration records for the participants were collected for up to four years after the program, but results were mixed. While participation in the program decreased the probability that offenders would be rearrested for drug-related offenses, it did not decrease the likelihood that they would be rearrested for other offenses.

### Before-and-After Designs

The common feature of before-and-after designs is the absence of a comparison group: All cases are exposed to the experimental treatment. The basis for comparison is instead provided by the pretreatment measures in the experimental group. These designs are thus useful for studies of interventions that are experienced by virtually every case in some population.

The simplest type of before-and-after design is the fixed-sample panel design (panel study). In a fixed-sample panel design, the same individuals are studied over time; the research may entail one pretest and one posttest. However, this type of before-and-after design does not qualify as a quasi-experimental design because comparing subjects with themselves at just one earlier point in time does not provide an adequate comparison group. Many influences other than the experimental treatment may affect a subject following the pretest—for instance, basic life experiences for a young subject.

A more powerful way to ensure that the independent variable actually affected the dependent variable when using a before-and-after design is by using a multiple-group before-and-after design. In this design, before-and-after comparisons are made of the same variables between different groups. Time series designs, sometimes called repeated measures panel designs, include several pretest

### Exhibit 6.8 Quasi-Experimental Design of Drug Court Research

<table>
<thead>
<tr>
<th>Drug-involved offenders</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>Participated in drug court</td>
<td>Rearrested</td>
</tr>
<tr>
<td>Control group</td>
<td>Did not participate</td>
<td>Rearrested</td>
</tr>
</tbody>
</table>

and posttest observations, allowing the researcher to study the process by which an intervention or treatment has an impact over time; hence, they are better than a simple before-and-after study.

These designs are particularly useful for studying the impact of new laws or social programs that affect large numbers of people and that are readily assessed by some ongoing measurement. For example, we might use a time series design to study the impact of a new seat belt law on the severity of injuries in automobile accidents using a monthly state government report on insurance claims. Specific statistical methods are required to analyze time series data, but the basic idea is simple: Identify a trend in the dependent variable up to the date of the intervention, and then control for outside influences and project the trend into the post-intervention period. This projected trend is then compared to the actual trend of the dependent variable after the intervention. A substantial disparity between the actual and projected trend is evidence that the intervention or event had an impact (Rossi & Freeman, 1989).

How well do these before-and-after designs meet the five criteria for establishing causality? The before–after comparison enables us to determine whether an association exists between the intervention and the dependent variable (because we can determine whether there was a change after the intervention). They also clarify whether the change in the dependent variable occurred after the intervention, so time order is not a problem. However, there is no control group, so we cannot rule out the influence of extraneous factors as the actual cause of the change we observe; spuriousness may be a problem. Some other event may have occurred during the study that resulted in a change in posttest scores. Overall, the longitudinal nature (the measurement of a phenomenon over a long period of time) of before-and-after designs can help to identify causal mechanisms, while the loosening of randomization requirements makes it easier to conduct studies in natural settings, where we learn about the influence of contextual factors. We will discuss the element of time in research in greater detail later in the chapter.

Case Study

The Effects of the Youth Criminal Justice Act

Carrington and Schulenberg’s (2008) study of the effect of the Youth Criminal Justice Act (YCJA) of 2002 in Canada on police discretion with apprehended young offenders illustrates a time series design. This design typically includes many pretest and posttest observations that allow the researcher to study the process by which an intervention or a treatment has an impact over time.

One of the major objectives of the YCJA, which came into effect in 2003 in Canada, was to reduce the number of referrals to youth court. The YCJA generally requires police officers who are thinking of charging a minor with a crime to first consider extralegal judicial measures such as giving the youth an informal warning.

To study the effects of the YCJA, Carrington and Schulenberg (2008) examined the number of juveniles who were apprehended and charged from January 1, 1986, through December 31, 2006. The Canadian Uniform Crime Reporting Survey records the number of minors who were charged as well as the number who were chargeable but not charged. The researchers note, “A change in the charge ratio, or proportion of chargeable youth who were charged, is an indication of a change in the use of police discretion with apprehended youth” (p. 355). To control for the actual crime rate of youth, the researchers also examined per capita ratios. Exhibit 6.9 displays the annual rates per 100,000 young persons who were (a) apprehended (i.e., chargeable), (b) charged, and (c) not charged. This clearly shows that the YCJA may have had the intended effect. Of course, the study design leaves open the possibility that something else in 2003 may have happened to effect this change in formal charges against juveniles. However, because there was no known event that could have had such a national impact, the conclusion that this effect is attributable to the YCJA is more plausible. As you can see, this time series design is particularly useful for studying the impact of new laws or social programs that affect everyone and can be readily assessed by ongoing measurement.

Ex Post Facto Control Group Designs

The ex post facto control group design appears to be very similar to the nonequivalent control group design and is often confused with it, but it does not meet as well the criteria for quasi-experimental designs. Similar to nonequivalent control group designs, this design has experimental and comparison groups that are not created by random assignment.
However, unlike the groups in nonequivalent control group designs, the groups in ex post facto designs are designated after the treatment has occurred. The problem with this is that if the treatment takes any time at all, people with particular characteristics may select themselves for the treatment or avoid it. Of course, this makes it difficult to determine whether an association between group membership and outcome is spurious. However, the particulars will vary from study to study; in some circumstances, we may conclude that the treatment and control groups are so similar that causal effects can be tested (Rossi & Freeman, 1989).

### Case Study

**Does an Arrest Increase Delinquency?**

David P. Farrington's (1977) classic study of how arrest sometimes increases delinquency, called the *deviance amplification process*, is an excellent example of an ex post facto control group design (Exhibit 6.10). Farrington tested the hypothesis that juveniles who were publicly labeled as deviant through being convicted of a delinquent act would increase their deviant behavior compared with those who were not so labeled. Using secondary data from the Cambridge Study of Delinquent Development, Farrington measured outcomes of 400 London working-class youths from age 8 to 18. Results indicated that youth who were labeled as delinquent (through conviction) subsequently committed more delinquent acts than similar youth who were not labeled in this way.

### What Are the Threats to Internal Validity and Generalizability in Experiments?

Like any research design, experimental designs must be evaluated for their ability to yield valid conclusions. True experiments are particularly well suited to producing valid conclusions about causality (internal validity), but they...
less likely to fare well in achieving generalizability. Quasi-experiments may provide more generalizable results than true experiments, but they are more prone to problems of internal invalidity (although some design schemes allow the researcher to rule out almost as many potential sources of internal invalidity as does a true experiment). In general, non-experimental designs (such as survey research and participant observation) permit less certainty about internal validity.

### Causal (Internal) Validity

An experiment’s ability to yield valid conclusions about causal effects is determined by the comparability of its experimental and comparison groups. First, of course, a comparison group must be created. Second, this comparison group must be so similar to the experimental group that it will show what the experimental group would be like if it did not receive the experimental treatment—that is, if the independent variable was not varied. For example, the only difference between the two groups in Bushman’s (2012) study was that one group watched a violent movie and the other group did not.

There are five basic sources of internal invalidity:

1. **Selection bias**—when characteristics of the experimental and comparison group subjects differ
2. **Endogenous change**—when the subjects develop or change during the experiment as part of an ongoing process independent of the experimental treatment
3. **External events/history effects**—when something occurs during the experiment, other than the treatment, that influences outcome scores
4. **Contamination**—when either the experimental group or the comparison group is aware of the other group and is influenced in the posttest as a result (Mohr, 1992)
5. **Treatment misidentification**—when variation in the independent variable (the treatment) is associated with variation in the observed outcome, but the change occurs through a process that the researcher has not identified

#### Selection Bias

You may already realize that the composition of the experimental and comparison groups in a true experiment is unlikely to be affected by their difference. If it were affected, it would cause selection bias. Random assignment equates the groups’
characteristics, though with some possibility for error due to chance. The likelihood of difference due to chance can be identified with appropriate statistics.

Even when the random assignment plan works, the groups can differ over time because of differential attrition, or what can be thought of as deselection—that is, the groups become different because, for various reasons, some subjects drop out of groups. This is not a likely problem for a laboratory experiment that occurs in one session, but for experiments in which subjects must participate over time, differential attrition may become a problem.

When subjects are not randomly assigned to treatment and comparison groups, as in nonequivalent control group designs, there is a serious threat of selection bias. Even if the researcher selects a comparison group that matches the treatment group on important variables, there is no guarantee that the groups were similar initially in terms of the dependent variable or another characteristic that ultimately influences posttest scores.

**Endogenous Change**

The type of problem considered an endogenous change occurs when natural developments in the subjects, independent of the experimental treatment, account for some or all of the observed change between pretest and posttest. Endogenous change includes three specific threats to internal validity:

1. **Testing.** Taking the pretest can in itself influence posttest scores. Subjects may learn something or be sensitized to an issue by the pretest and, as a result, respond differently when they are asked the same questions in the posttest.

2. **Maturation.** Changes in outcome scores during experiments that involve a lengthy treatment period may be due to maturation. Subjects may age or gain experience in school or grow in knowledge, all as part of a natural maturation experience, and thus respond differently on the posttest from the way they responded on the pretest.

3. **Regression.** People experience cyclical or episodic changes that result in different posttest scores, a phenomenon known as a regression effect. Subjects who are chosen for a study because they received very low scores on a test may show improvement in the posttest, on average, simply because some of the low scorers had been having a bad day. It is hard, in many cases, to know whether a phenomenon is subject to naturally occurring fluctuations, so the possibility of regression effects should be considered whenever subjects are selected because of their initial extremely high or low values on the outcome variable (Mohr, 1992).

Testing, maturation, and regression effects are generally not a problem in true experiments. Both the experimental group and the comparison group take the pretest, and they are both subject to maturation and regression effects, so even if these possibilities lead to a change in posttest scores, the comparison between the experimental and control groups will not be affected because the groups started off with similar characteristics. Of course, in experiments with no pretest, testing effects themselves are not a problem. However, in most before-and-after designs without a comparison group, endogenous change effects could occur and lead to an invalid conclusion that there had been an effect of the independent variable.

**External Events**

External events (sometimes referred to as the history effect during the experiment)—things that happen outside the experiment—can also change the subjects’ outcome scores. An example of this is a newsworthy event that is relevant to the focus of an experiment to which subjects are exposed. What if researchers were evaluating the effectiveness of a mandatory arrest policy in decreasing incidents of intimate partner assault and an event such as the
murder trial of O. J. Simpson occurred during the experiment? This would clearly be a historical event that might compromise the results. This trial saw a momentous amount of media coverage, and as a result, intimate partner assault and homicide were given a tremendous amount of attention. Because of this increased awareness, many victims of intimate partner violence reported their victimizations during this time—police agencies and women’s shelters were flooded with calls. If a researcher had been using calls to police in a particular jurisdiction as an indicator of the incidence of intimate partner assault, this historical event would have seriously jeopardized the internal validity of his or her results. Why? Because the increase in police calls would have had more to do with the trial than with any recent change in arrest policies.

**Contamination**

Contamination occurs in an experiment when the comparison group is in some way affected by, or affects, the treatment group. This problem basically arises from failure to adequately control the conditions of the experiment. When comparison group members become aware that they are being denied some advantage, they may increase their efforts to compensate, creating a problem called compensatory rivalry or the John Henry effect (Cook & Campbell, 1979). On the other hand, control group members may become demoralized if they feel that they have been left out of some valuable treatment and may perform worse than they would have outside the experiment. The treatment may seem, in comparison, to have a more beneficial effect than it actually did. Both compensatory rivalry and demoralization can thus distort the impact of the experimental treatment.

The danger of contamination can be minimized if the experiment is conducted in a laboratory, if members of the experimental group and the comparison group have no contact while the study is in progress, and if the treatment is relatively brief. To the degree that these conditions are not met, the likelihood of contamination will increase.

**Treatment Misidentification**

Treatment misidentification occurs when subjects experience a treatment that wasn’t intended by the researcher. Treatment misidentification has at least three sources:

1. **Expectancies of experimental staff.** Change among experimental subjects may be due to the positive expectancies of the experimental staff who are delivering the treatment rather than due to the treatment itself. Such positive staff expectations can create a self-fulfilling prophecy and can occur even in randomized experiments when well-trained staff convey their enthusiasm for an experimental program to the subjects in subtle ways. These expectancy effects can be very difficult to control in field experiments. However, in some experiments concerning the effects of treatments such as medical drugs, double-blind procedures can be used. Staff will deliver the treatments without knowing which subjects are getting the treatment and which are receiving a placebo—something that looks similar to the treatment but has no effect. In fact, the prison experiment discussed earlier in this chapter used a double-blind procedure to randomly assign inmates to a security classification category. In the experiment, only the executive director of the corrections department and the director of classification were aware of the research. Correctional staff, other individuals who worked with the inmates, and the inmates themselves were unaware of the study. In this way, any expectancies that the staff may have had were unlikely to affect inmate behavior.

---

**Contamination** A source of causal invalidity that occurs when the experimental and/or the comparison group is aware of the other group and is influenced in the posttest as a result.

**Compensatory rivalry (John Henry effect)** A type of contamination in experimental and quasi-experimental designs that occurs when control group members are aware that they are being denied some advantage and increase their efforts by way of compensation.

**Treatment misidentification** A problem that occurs in an experiment when the treatment itself is not what causes the outcome, but rather the outcome is caused by some intervening process that the researcher has not identified and is not aware of.

**Expectancies of the experimental staff (self-fulfilling prophecy)** A source of treatment misidentification in experiments and quasi-experiments that occurs when change among experimental subjects is due to the positive expectancies of the staff who are delivering the treatment rather than to the treatment itself.

**Double-blind procedure** An experimental method in which neither subjects nor the staff delivering experimental treatments know which subjects are getting the treatment and which are receiving a placebo.
2. Placebo effect. Treatment misidentification may occur when subjects receive a treatment that they consider likely to be beneficial and then improve because of that expectation rather than the treatment itself. In medical research, the placebo is often a chemically inert substance that looks similar to the experimental drug but actually has no medical effect. Research indicates that the placebo effect produces positive health effects in two thirds of patients suffering from relatively mild medical problems (Goleman, 1993). Placebo effects can also occur in social science research. The only way to reduce this threat to internal validity is to treat the comparison group with something similar.

3. Hawthorne effect. Members of the treatment group may change in terms of the dependent variable because their participation in the study makes them feel special. This problem can occur when treatment group members compare their situation to that of the control group members who are not receiving the treatment. In this case, this is a type of contamination effect. However, experimental group members could feel special simply because they are in the experiment. The Hawthorne effect is named after a famous productivity experiment at the Hawthorne electric plant outside Chicago. Workers were moved to a special room for a study of the effects of lighting intensity and other work conditions on their productivity. After this move, the workers began to increase their output no matter what change was made in their working conditions, even when the conditions became worse. The researchers concluded that the workers felt they should work harder because they were part of a special experiment.

### Interaction of Testing and Treatment

A variation of the problem of external validity occurs when the experimental treatment is effective only when particular conditions created by the experiment occur. For example, if subjects have had a pretest, it may sensitize them to a particular issue, so when they are exposed to the treatment, their reaction is different from what it would have been if they had not taken the pretest. In other words, testing and treatment interact to produce the outcome.

Suppose you were interested in the effects of a diversity training film on prejudicial attitudes. After answering questions in a pretest about their attitudes on various topics related to diversity (e.g., racial or sexual prejudice), the subjects generally became more sensitive to the issue of prejudice without seeing the training film. On the posttest, then, their attitudes may be different from pretest attitudes simply because they have become sensitized to the issue of diversity through pretesting. In this situation, the treatment may actually have an effect, but it would be difficult to determine how much of the effect was attributable to the sensitizing pretest and how much was due to seeing the film.

This possibility can be tested with what is called the Solomon four-group design. In this version of a true experimental design, subjects are randomly assigned to at least two experimental groups and at least two comparison groups. One

### Exhibit 6.11 Solomon Four-Group Design Testing the Interaction of Pretesting and Treatment

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>R</th>
<th>O1</th>
<th>X</th>
<th>O2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison group</td>
<td>R</td>
<td>O1</td>
<td>X</td>
<td>O2</td>
</tr>
<tr>
<td>Experimental group</td>
<td>R</td>
<td>X</td>
<td></td>
<td>O2</td>
</tr>
<tr>
<td>Comparison group</td>
<td>R</td>
<td></td>
<td>O2</td>
<td></td>
</tr>
</tbody>
</table>

Note: R = random assignment; O = observation (pretest or posttest); X = experimental treatment.
experimental group and one comparison group will have a pretest, and the other two groups will not have a pretest (see Exhibit 6.11). If testing and treatment do interact, the difference in outcome scores between the experimental and comparison groups will differ between the subjects who took the pretest and those who did not.

Ultimately, the external validity of experimental results will increase with the success of replications taking place at different times and places, using different forms of the treatment. As indicated by the replications of the Sherman and Berk (1984) study of arrest for domestic violence, the result may be a more detailed, nuanced understanding of the hypothesized effect.

**Generalizability**

The need for generalizable findings can be thought of as the Achilles heel of the true experimental design. The design components that are essential for a true experiment and minimize the threats to causal (internal) validity also make it more difficult to achieve sample generalizability, or the ability to apply the findings to a clearly defined, larger population.

**Sample Generalizability**

Subjects who can be recruited for a laboratory experiment, randomly assigned to a group, and kept under carefully controlled conditions for the study’s duration are often not a representative sample of any large population of interest. In fact, most are recruited from college populations. Can they be expected to react to the experimental treatment in the same way as members of the larger population who are not students or may never have gone to college? The more artificial the experimental arrangements, the greater the problem can be (Campbell & Stanley, 1996).

Not only do the characteristics of the subjects themselves determine the generalizability of the experimental results, but the generalizability of the treatment and of the setting for the experiment also must be considered (Cook & Campbell, 1979). Field experiments are likely to yield findings that are more generalizable to broader populations than are laboratory experiments using subjects who must volunteer. When random selection is not feasible, the researchers may be able to increase generalizability by selecting several sites for conducting the experiments that offer obvious contrasts in the key variables of the population. The follow-up studies to Sherman and Berk’s (1984) work, for example (see Chapter 2), were conducted in cities that differed from Minneapolis, the original site, in social class and ethnic composition. As a result, although the findings are not statistically generalizable to a larger population, they do give some indication of the study’s general applicability (Cook & Campbell, 1979).

### The Element of Time in Research

Nonexperimental research designs can be either cross-sectional or longitudinal. In a cross-sectional research design, all data are collected at one point in time. Identifying the time order of effects—what happened first, second, and so on—is critical for developing a causal analysis but can be an insurmountable problem with a cross-sectional design. In longitudinal research designs, data are collected at two or more points in time, so identification of the time order of effects can be quite straightforward. You can think of an experiment as a type of longitudinal design because subjects are often observed at two or more points in time.

Much of the research you have encountered so far in this text has been cross-sectional. Although each survey and interview takes some time to carry out, if it measures the actions, attitudes, and characteristics of respondents at only one time, it is considered cross-sectional. The name comes from the idea that a snapshot from a cross-section of the population is obtained at one point in time.

---

**Sample generalizability** Exists when a conclusion based on a sample, or subset, of a larger population holds true for that population.

**Cross-sectional research design** A study in which data are collected at only one point in time.

**Longitudinal research design** A study in which data are collected that can be ordered in time; also defined as research in which data are collected at two or more points in time.
Lo, Kim, and Cheng (2008) provide an interesting example of an attempt to collect retrospective data using a cross-sectional design. The researchers wanted to determine if certain offenders were more likely to repeat the same crimes or commit different crimes. Specifically, they wanted to know if the type of crime committed early in someone’s life was a reliable predictor of offenses committed later. Stated differently, do offenders specialize in different types of crime? Lo et al. obtained official arrest records from the age of 18 to the time of the study (typically around age 25) for a sample of young offenders who were incarcerated in county jails in Ohio. They then asked the inmates to reconstruct their drug and alcohol use among other things monthly for the same time period using a life calendar instrument based on the Arrestee Drug Abuse Monitoring (ADAM) interview schedule. This life calendar instrument helps respondents recall events in their past by displaying each month of a given year along with key dates noted within the calendar, such as birthdays, arrests, holidays, anniversaries, and so on. Respondents are given a calendar that displays these key dates, typically called anchors, and then are asked to recall the variables of interest (i.e., drug use, victimizations) that also occurred during the specified time frame. The use of a life calendar has been shown to improve the ability of respondents to recall events in the past compared to basic questions without a calendar (Belli, Stafford, & Alwin, 2009).

Results of Lo et al.’s (2008) research were somewhat mixed regarding offender specialization. Most offenders engaged in a variety of offenses prior to their current arrest. However, compared to drug and property offenders, violent offenders were more likely to specialize, as they were the most likely to have had violent arrest records prior to their current offenses.

It is important to note that retrospective data such as these are often inadequate for measuring variation in past psychological states or behaviors, because what we recall about our feeling or actions in the past is likely to be influenced by...
what we feel in the present. For example, retrospective reports by both adult alcoholics and their parents appear to greatly overestimate the frequency of childhood problems (Vaillant, 1995). People cannot report reliably the frequency and timing of many past events, from hospitalization to hours worked. However, retrospective data tend to be reliable when they concern major, persistent experiences in the past, such as what type of school someone went to or how a person’s family was structured (Campbell, 1992).

In contrast, longitudinal research collects data at two or more points in time and, as such, data can be ordered in time. By measuring the value of cases on an independent variable and a dependent variable at different times, the researcher can determine whether variation in the independent variable precedes variation in the dependent variable.

In some longitudinal designs, the same sample (or panel) is followed over time. In other longitudinal designs, sample members are rotated or completely replaced. The population from which the sample is selected may be defined broadly, as when a longitudinal survey of the general population is conducted, or the population may be defined narrowly, as when members of a specific age group are sampled at multiple points in time. The frequency of follow-up measurement can vary, ranging from a before-and-after design with just one follow-up to studies in which various indicators are measured every month for many years.

Collecting data at two or more points in time rather than at one time can prove difficult for a number of reasons: lack of long-term funding, participant attrition, and so on. Quite frequently, researchers cannot or are simply unwilling to delay completion of a study for even one year in order to collect follow-up data. But think of the many research questions that really should involve a much longer follow-up period: Does community-oriented policing decrease rates of violent crime? What is the impact of job training in prison on recidivism rates? How effective are batterer treatment programs for individuals convicted of intimate partner assault? Do parenting programs for young mothers and fathers reduce the likelihood of their children becoming delinquent? It is safe to say that we will never have enough longitudinal data to answer many important research questions. Nonetheless, the value of longitudinal data is so great that every effort should be made to develop longitudinal research designs when they are appropriate for the research question being asked. The following discussion of the three major types of longitudinal designs will give you a sense of the possibilities (see Exhibit 6.12).

### Exhibit 6.12 Three Types of Longitudinal Design

```plaintext
Exhibit 6.12 Three Types of Longitudinal Design

Repeated Cross-Sectional Design (Trend Study)

Fixed-Sample Panel Design (Panel Study)

Event-Based Design (Cohort Study)
```

**Repeated Cross-Sectional Designs**

Studies that use a repeated cross-sectional design, also known as a trend study, have become fixtures of the political arena around election time. Particularly in presidential election years, we accustom ourselves to reading weekly, even daily, reports on the percentage of the population that supports each candidate. Similar polls are conducted to track sentiment on many other social issues. For example, a 1993 poll reported that 52% of adult Americans supported a ban on the possession of handguns compared with 41% in a similar poll conducted in 1991. According to pollster Louis Harris, this increase indicated a “sea change” in public attitudes...
Another researcher said, “It shows that people are responding to their experience [of an increase in handgun-related killings]” (cited in Barringer, 1993, p. A14).

Repeated cross-sectional surveys are conducted as follows:

1. A sample is drawn from a population at Time 1, and data are collected from the sample.
2. As time passes, some people leave the population and others enter it.
3. At Time 2, a different sample is drawn from this population.

Fixed-Sample Panel Designs

We talked about fixed-sample panel designs above when we highlighted quasi-experimental designs. However, these types of designs can also be used when no experimental treatment or program is being examined. Panel designs allow us to identify changes in individuals, groups, or whatever we are studying. This is the process for conducting fixed-sample panel designs:

1. A sample (called a panel) is drawn from a population at Time 1, and data are collected from the sample.
2. As time passes, some panel members become unavailable for follow-up, and the population changes.
3. At Time 2, data are collected from the same people as at Time 1 (the panel), except for those people who cannot be located.

Because a panel design follows the same individuals, it is better than a repeated cross-sectional design for testing causal hypotheses. For example, Sampson and Laub (1990) used a fixed-sample panel design to investigate the effect of childhood deviance on adult crime. They studied a sample of white males in Boston when the subjects were between 10 and 17 years old and then followed up when the subjects were in their adult years. Data were collected from multiple sources, including the subjects themselves and criminal justice records. The researchers found that children who had been committed to a correctional school for persistent delinquency were much more likely than other children in the study to commit crimes as adults: 61% were arrested between the ages of 25 and 32, compared with 14% of those who had not been in correctional schools as juveniles (p. 614). In this study, juvenile delinquency unquestionably occurred before adult criminality. If the researchers had used a cross-sectional design to study the past of adults, the juvenile delinquency measure might have been biased by memory lapses, by self-serving recollections about behavior as juveniles, or by loss of agency records. The problem, of course, is that tracking people for years is extremely expensive, and many people in the original sample drop out for various reasons. Panel designs are also a challenge to implement successfully, and often are not even attempted, because of two major difficulties:

1. Expense and attrition. It can be difficult, and very expensive, to keep track of individuals over a long period, and inevitably, the proportion of panel members who can be located for follow-up will decline over time. Panel studies often lose more than one quarter of their members through attrition (Miller, 1991, p. 170), and because those who are lost are often dissimilar to those who remain in the panel, the sample’s characteristics begin to change and internal validity is compromised.

2. Subject fatigue. Panel members may grow weary of repeated interviews and drop out of the study, or they may become so used to answering the standard questions in the survey that they start giving stock answers rather than actually thinking about their current feelings or actions (Campbell, 1992). This is called the problem of subject fatigue. Fortunately, subjects do not often seem to become fatigued in this way, particularly if the research staff have maintained positive relations with them.

Repeated cross-sectional design (trend study) A type of longitudinal study in which data are collected at two or more points in time from different samples of the same population

Subject fatigue Problems caused by panel members growing weary of repeated interviews and dropping out of a study or becoming so used to answering the standard questions in the survey that they start giving stock or thoughtless answers
Event-Based Designs

In an event-based design, often called a cohort study, the follow-up samples (at one or more times) are selected from the same cohort, people who all have experienced a similar event or a common starting point. Examples include the following:

- **Birth cohorts:** those who share a common period of birth (those born in the 1940s, 1950s, 1960s, etc.)
- **Seniority cohorts:** those who have worked at the same place for about five years, about 10 years, and so on
- **School cohorts:** freshmen, sophomores, juniors, and seniors

### Causality in Nonexperimental Designs

How well do the research designs just described satisfy the criteria necessary to determine causality? Although it is relatively easy to establish that an empirical association exists between an independent and a dependent variable in these designs, the other criteria are much more difficult to assess.

Let us first illustrate the importance of time order and nonspuriousness using research that has examined the factors related to the gender and crime relationship. Based on both victimization data and official police reports, data indicate that males commit the majority of all crimes. Why is this? Gottfredson and Hirschi’s (1990) general theory of crime (GTC) contends that the reason males engage in more criminality is that they have lower levels of self-control than females. The researchers also contend that socialization of children by parents is the primary factor in the development of self-control. However, based on a critique of the GTC by Miller and Burack (1993) and the power control theory (Hagan, Gillis, & Simpson, 1985), Blackwell and Piquero (2005) hypothesized that the power relationships that exist between parents in a household (e.g., patriarchy) would also affect the socialization experiences of boys and girls and, ultimately, their levels of self-control. To summarize briefly, Blackwell and Piquero examined the factors related to self-control acquisition in childhood using a sample of adults. Using this same sample of adults, they then examined the extent to which low self-control predicted the propensity for criminal offending. In a nutshell, they sought to explain the origins of self-control as well as the effects of self-control on criminal offending and how all this may be different for males and females from patriarchal families and for males and females from more egalitarian families. Using a random sample of 350 adults from Oklahoma City, Oklahoma, they found that there were indeed differences in the way power relationships between parents affected the acquisition of self-control for males and females. They also found, however, that there were essentially no differences in the ability of self-control to predict criminal aspirations; males and females with low self-control were more likely to self-report that they would engage in criminal behavior than their counterparts with higher self-control.

Do these findings establish that low self-control leads to crime through poor socialization of children by parents? Well, there are many assumptions being made here that we hope you can see right away. First, this study relied on the recollections of adults about their childhood socialization. It also assumed that levels of low self-control were subsequent to parental socialization and preceded individuals’ aspirations to offend (time order). This may very well be the case. It may be that those adults who were more likely to offend had inadequate socialization, which created low self-control. However, it may be that offending behavior during their adolescence led to weak attachments to family and high attachments to other delinquent peers similar to themselves, which also decreased levels of self-control. In this case, the delinquent offending and peer associations would be a third variable responsible for both the low self-control and the criminal aspirations in adulthood (spurious relationship). The problem, of course, is that with cross-sectional data such as these, the correct time order cannot be established and it is difficult to control for the effects of all important factors. Blackwell and Piquero (2005) stated this limitation well when they concluded, “Future research should attempt to examine the changing nature of parental socialization and self-control across gender in longitudinal studies” (p. 15).
To reduce the risk of spuriousness, Blackwell and Piquero (2005) used the technique of statistical control. Exhibit 6.13 represents the important concept of statistical control with a hypothetical study of the relationship between attending a boot camp in prison (a highly regimented, discipline-focused rehabilitation program) and the likelihood of committing crimes after prison (the recidivism rate). In Exhibit 6.12, the data for all prisoners show that prisoners who attended boot camp were less likely to return to committing crimes after they left prison. However, as the more detailed data show, more female prisoners attended boot camp than male prisoners, so gender may have played a significant role in recidivism. The researchers, however, reduced the risk of spuriousness by using two statistical control methods: They examined the association between attending boot camp and post-prison criminality for men and for women. After doing this, researchers determined that boot camp did not reduce recidivism. It just appeared to do so, because women were more likely to attend boot camp and less likely to commit crimes after prison, regardless of whether they attended boot camp.

Our confidence in causal conclusions based on nonexperimental research also increases with identification of a causal mechanism. These mechanisms are called intervening variables in nonexperimental research and help us understand how variation in the independent variable results in variation in the dependent variable. Of course, identification of one (or two or three) intervening variables does not end the possibilities for clarifying the causal mechanisms. You might ask why structural disadvantage tends to result in lower levels of family social control or how family social control influences delinquency. You could then conduct research to identify the mechanisms that link, for example, family social control and juvenile delinquency. (Perhaps the children feel they are not cared for, so they become less concerned with conforming to social expectations.) This process could go on and on. The point is that identification of a mechanism through which the independent variable influences the dependent variable increases our confidence in the conclusion that a causal connection does indeed exist.

When you think about the role of variables in causal relationships, do not confuse variables that cause spurious relationships with variables that intervene in causal relationships, even though both are third variables that do not appear in the initial hypothesis. Intervening variables help explain the relationship between the independent variable (juvenile delinquency) and the dependent variable (adult criminality).

Nonexperimental research can be a very effective tool for exploring the context in which causal effects occur. Administering surveys in many different settings and to different types of individuals is usually much easier than administering various experiments. The difficulty of establishing nonspuriousness does not rule out using nonexperimental data to evaluate causal hypotheses. In fact, when enough nonexperimental data are collected to allow tests of multiple implications of the same causal hypothesis, the results can be very convincing (Rossi & Freeman, 1989).

In any case, nonexperimental tests of causal hypotheses will continue to be popular because the practical and ethical problems in randomly assigning people to different conditions preclude the test of many important hypotheses with an experimental design. Just remember to carefully consider possible sources of spuriousness and other problems when evaluating causal claims based on individual nonexperimental studies.

How Do Experimenters Protect Their Subjects?

Social science experiments often raise difficult ethical issues. You have already read in Chapter 3 about Philip Zimbardo’s (2004) Stanford Prison Experiment. This experiment was actually ended after only six days, rather than after the planned two weeks, because of the psychological harm that seemed to result from the unexpectedly sadistic behavior of some of the “guards.” Although Zimbardo’s follow-up research convinced him that there had been no lasting harm to subjects, concern about the potential for harm would preclude many such experiments today.

**Statistical control** A technique used in nonexperimental research to reduce the risk of spuriousness. One variable is held constant so the relationship between two or more other variables can be assessed without the influence of variation in the control variable.

**Intervening variables** Variables that are influenced by an independent variable and in turn influence variation in a dependent variable, thus helping to explain the relationship between the independent and dependent variables.
In spite of the ethical standard of disclosure and informed consent by subjects, deception is an essential part of many experimental designs. As a result, contentious debate continues about the interpretation of this standard. Experimental evaluation of social programs also poses ethical dilemmas because they require researchers to withhold possibly beneficial treatment from some of the subjects solely on the basis of chance (Boruch, 1997). In this section, we will give special attention to the problems of deception and the distribution of benefits in experimental research.

**Deception**

Deception is used in social experiments to create more “realistic” treatments, often within the confines of a laboratory. You learned in Chapter 3 about Stanley Milgram’s (1965) use of deception in his classic study of obedience to authority. Volunteers were recruited for what they were told was a study of the learning process, not a study of obedience to authority. The experimenter told the volunteers that they were administering electric shocks to a “student” in the next room, when there were actually neither students nor shocks. Subjects seemed to believe the deception.

Whether or not you believe that you could be deceived in this way, you are not likely to be invited to participate in an experiment such as Milgram’s. Current federal regulations preclude deception in research that might trigger such upsetting feelings. However, deception is still routine in many college laboratories. The question that must always be answered is, “Is there sufficient justification to allow the use of deception?” David Willer and
Henry A. Walker (2007) pay particular attention to debriefing after deception in their book about experimental research. They argue that every experiment involving deception should be followed immediately for each participant with debriefing, sometimes called dehoaxing, in which the deception is explained, and all the participants’ questions are answered to their satisfaction; those participants who still feel aggrieved are directed to a university authority to file a complaint or to a counselor for help with their feelings. This is sound advice.

**Selective Distribution of Benefits**

Field experiments conducted to evaluate social programs also can involve issues of informed consent (Hunt, 1985). One ethical issue that is somewhat unique to field experiments is the selective distribution of benefits: How much are subjects harmed by the way treatments are distributed in the experiment? For example, Sherman and Berk’s (1984) experiment, and its successors, required police to make arrests in domestic violence cases largely on the basis of a random process. When arrests were not made, did the subjects’ abused spouses suffer?

Is it ethical to give some potentially advantageous or disadvantageous treatment to people on a random basis? For example, in the drug court field experiment, is it ethical to randomly assign those who wanted extra help with their drug problem to the comparison group that did not receive extra treatment? Random distribution of benefits is justified when the researchers do not know whether some treatment actually is beneficial—and, of course, it is the goal of the experiment to find out. Chance is as reasonable a basis for distributing the treatment as any other. Also, if insufficient resources are available to fully fund a benefit for every eligible person, distribution of the benefit on the basis of chance to equally needy persons is ethically defensible (Boruch, 1997).

### CAREERS AND RESEARCH

**Amanda Aykanian, Research Associate, Advocates for Human Potential**

Amanda Aykanian majored in psychology at Framingham State University and found that she enjoyed the routine and organization of research. She wrote an undergrad thesis to answer the research question: How does the way in which course content is presented affect students’ feelings about the content and the rate at which they retain it?

After graduating, Aykanian didn’t want to go to graduate school right away; instead she wanted to explore her interests and get a sense of what she could do with research. Advocates for Human Potential (AHP) was the last research assistant (RA) job that Aykanian applied for. Her initial tasks as an RA at AHP ranged from taking notes, writing agendas, and assembling project materials to entering research data, cleaning data, and proofing reports. As she contributed more to project reports, she began to think about data from a more theoretical standpoint.

During 7 years at AHP, Aykanian has helped lead program evaluation research, design surveys and write survey questions, conduct phone and qualitative interviews, and lead focus groups. Her program evaluation research almost always uses a mixed-methods approach, so Aykanian has learned a lot about how qualitative and quantitative methods can complement each other. She has received a lot of on-the-job training in data analysis and has learned how to think about and write a proposal in response to federal funding opportunities.

Aykanian was promoted to research associate and describes her current role as part program evaluation coordinator and part data analyst. She has also returned to graduate school, earning a master’s degree in applied sociology and then starting a PhD program in social welfare.
Conclusion

True experiments play two critical roles in criminological research. First, they are the best research design for testing causal hypotheses. Even when conditions preclude the use of a true experimental design, many research designs can be improved by adding experimental components. Second, true experiments provide a comparison point for evaluating the ability of the other research designs to achieve causally valid results.

In spite of their obvious strengths, true experiments are used infrequently to study many research problems related to criminology and criminal justice. There are three basic reasons for this:

1. The experiments required to test many important hypotheses require more resources than most social scientists can access.
2. Most research problems of interest to social scientists simply are not amenable to experimental designs for reasons ranging from ethical considerations to the limited possibilities for randomly assigning people to different conditions in the real world.
3. The requirements of experimental design usually preclude large-scale studies and so limit generalizability to a degree that is unacceptable to many social scientists.

When a true experimental design is not feasible, researchers may instead use a quasi-experimental design, including nonequivalent control group designs, before-and-after designs, and ex post facto control group designs. As the studies highlighted in this chapter show, researchers examining issues related to criminology and criminal justice have been very creative in developing experimental research projects in the real world that can appropriately meet the demands of causal inference.

Key Terms

Anchors 141
Arrestee Drug Abuse Monitoring (ADAM) 141
Association 123
Before-and-after designs 132
Causal effect (idiographic perspective) 121
Ceteris paribus 120
Cohort 144
Compensatory rivalry (John Henry effect) 138
Contamination 138
Context 122
Contextual effect 126
Control or comparison group 127
Counterfactual 120
Cross-sectional research design 140
Debriefing 147
Differential attrition 137
Double-blind procedures 138
Endogenous change 137
Event-based design 144
Ex post facto control group design 132
Expectancies of the experimental staff (self-fulfilling prophecy) 138
Experimental group 127
External events (history effect) 137
Field experiment 130
Fixed-sample panel design (panel study) 133
Hawthorne effect 139
Idiographic causal explanation 121
Intervening variables 145
Life calendar 141
Longitudinal research design 140
Matching 129
Mechanism 125
NOMothetic causal explanation 120
Nonequivalent control group designs 132
Nonsurplusness 125
Placebo effect 139
Posttest 129
Pretest 129
Quasi-experimental design 131
Random assignment (randomization) 128
Regression effect 137
Repeated cross-sectional design (trend study) 142
Sample generalizability 140
Selection bias 136
Selective distribution of benefits 147
Solomon four-group design 139
Spurious 125
Statistical control 145
Subject fatigue 143
Time order 124
Time series design (repeated measures panel design) 133
Treatment misidentification 138
True experiment 127
Chapter 6  Causation and Experimentation

Highlights

- A causal explanation relies on a comparison. The value of cases on the dependent variable is measured after they have been exposed to variation on an independent variable. This measurement is compared to what the value of cases on the dependent variable would have been if they had not been exposed to the variation in the independent variable. The validity of causal conclusions rests on how closely the comparison group comes to the ideal counterfactual.

- Three criteria are generally viewed as necessary for identifying a causal relationship: association between the variables, proper time order, and nonspuriousness of the association. In addition, the basis for concluding that a causal relationship exists is strengthened by identification of a causal mechanism and the context for the relationship.

- Association between two variables is in itself insufficient evidence of a causal relationship. This point is commonly made with the expression, "Correlation does not prove causation."

- Experiments use random assignment to make comparison groups as similar as possible at the outset of an experiment in order to reduce the risk of spurious effects due to extraneous variables.

- Nonexperimental designs use statistical controls to reduce the risk of spuriousness. A variable is controlled when it is held constant so that the association between the independent and dependent variables can be assessed without being influenced by the control variable.

- Longitudinal designs are usually preferable to cross-sectional designs for establishing the time order of effects. Longitudinal designs vary in terms of whether the same people are measured at different times, how the population of interests is defined, and how frequently follow-up measurements are taken. Fixed-sample panel designs provide the strongest test for the time order of effects, but they can be difficult to carry out successfully because of their expense as well as subject attrition and fatigue.

- Ethical and practical constraints often prevent the use of experimental designs.

Exercises

Discussing Research

1. Review articles in several newspapers, copying down all causal assertions. These might include assertions that the presence of community policing was related positively to decreasing rates of violence, claims that the stock market declined because of uncertainty in the Middle East, or explanations about why a murder was committed. Inspect the articles carefully, noting all the evidence used to support the causal assertions. Which criteria for establishing causality are met? What other potentially important influences on the reported outcome have been overlooked?

2. Select several research articles in professional journals that assert or imply that they have identified a causal relationship between two or more variables. Is each of the criteria for establishing the existence of a causal relationship met? Find a study in which subjects were assigned randomly to experimental and comparison groups to reduce the risk of spurious influences on the supposedly causal relationship. How convinced are you by the study?

3. The web-based interactive exercises contain lessons on causation and experimental design. Try them out at this point.

Finding Research on the Web

1. Read an original article describing a social experiment. (Social psychology "readers," collections of such articles for undergraduates, are a good place to find interesting studies.) Critique the article, focusing on the extent to which experimental conditions were controlled and the causal mechanism was identified. Based on the study's control over conditions and identification of the causal mechanism, how confident were you in the causal conclusions?

2. Go to the website of the U.S. Department of Justice and examine its handbook on Community Policing (http://www.cops.usdoj.gov/pdf/vets-to-cops/e030917193-CP-Defined.pdf). What causal assertions are made on the site? Pick one of these assertions and propose a research design with which to test this assertion. Be specific.
3. Go to SocioSite (http://www.sociosite.net). Choose “Subject Areas.” Then choose “Crime” or “Criminology.” Find an example of research that has been done using experimental methods. Explain the experiment. Choose at least five of the key terms for this chapter (listed above) that are relevant to and incorporated in the research experiment you have located on the web. Explain how each of the five key terms you have chosen plays a role in the research example you have found.

4. Go to the Disaster Center website (http://www.disastercenter.com/crime/). Review the crime rate nationally and, by picking out links to state reports, compare the recent crime rates in two states. Report on the prevalence of the crimes you have examined. Propose a causal explanation for variation in crime between states, over time, or both. What research design would you propose to test this explanation? Explain.

Critiquing Research

1. Go to this book’s Study Site, https://study.sagepub.com/bach mainfrcjrs, and choose two research articles that include some attention to causality (as indicated by a check in that column of the article matrix). Describe the approach taken in each article to establish causality. How do the approaches differ from each other? Which approach seems stronger to you?

2. Select a true experiment, perhaps from the *Journal of Experimental Criminology* or from sources suggested in class. Diagram the experiment using the exhibits in this chapter as a model. Discuss the extent to which experimental conditions were controlled and the causal mechanism was identified. How confident can you be in the causal conclusions from the study, based on review of the threats to internal validity discussed in this chapter: selection bias, endogenous change, external events, contamination, and treatment misidentification? How generalizable do you think the study’s results are to the population from which the cases were selected? How generalizable are they to specific subgroups in the study? How thoroughly do the researchers discuss these issues?

3. Repeat Question 2 above with a quasi-experiment.

Making Research Ethical

1. Under what conditions do you think that randomized assignment of subjects to a specific treatment is ethical in criminal justice research? Was it ethical for the researchers who conducted experiments on the effect of arrest in intimate partner assault (Chapter 2) to randomly assign individuals accused of domestic violence to an arrest or nonarrest treatment? What about in a laboratory study with students such as yourself? Do you think it would be ethical to assign students randomly to different groups, with some receiving stressful stimuli such as loud noises?

2. Critique the ethics of one of the experiments presented in this chapter or some other experiment you have read about. What specific rules do you think should guide researchers’ decisions about subject deception and the selective distribution of benefits?

3. Lo et al. (2008) surveyed inmates in county jails. Federal regulations require special safeguards for research on prisoners. Do you think special safeguards are necessary? Why or why not? What type of research would you allow with prisoners: experiments, surveys, or observational studies? Do you think it is possible for prisoners to give voluntary consent to research participation? What procedures might help make prisoners’ consent to research truly voluntary?

4. Bushman and Huesmann (2012) tested the impact of watching a violent video on students’ level of aggressiveness. They found that watching the violent video increased aggressiveness. Do you consider it ethical to expose subjects to an intervention that might increase their aggressiveness? Are there any situations in which you would not approve of such research? Any types of subjects you would exclude from such research? Any limits you would draw on the type of intervention that could be tested? Would you impose any requirements for debriefing?

Developing a Research Proposal

How will you try to establish the causal effects that you hypothesize?

1. Identify at least one hypothesis involving what you expect is a causal relationship.
2. Identify key variables that should be controlled in your survey design in order to decrease the possibility of arriving at a spurious conclusion about the hypothesized causal effect. Draw on relevant research literature and social theory to identify these variables.

3. Add a longitudinal component to your research design. Explain why you decided to use this particular longitudinal design.

4. Review the criteria for establishing a causal effect and discuss your ability to satisfy each one.

Performing Data Analysis in SPSS or Excel

<table>
<thead>
<tr>
<th>Data for Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dataset</strong></td>
</tr>
<tr>
<td>Monitoring the</td>
</tr>
<tr>
<td>future 2013 grade</td>
</tr>
<tr>
<td>10.sav</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables for Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable Name</strong></td>
</tr>
<tr>
<td>Lowparenteduc</td>
</tr>
<tr>
<td>V7234</td>
</tr>
<tr>
<td>White</td>
</tr>
</tbody>
</table>

1. Take a look at some of the variables in this dataset. Is the MTF dataset suited to nomothetic or idiographic explanations? Why?

2. The link between a parent’s education and a child’s subsequent outcomes is well documented in many different fields, from criminology to health. Let’s look at the relationship between a parent’s education level and whether a respondent has ever been suspended from school.

   a. State a hypothesis about the relationship between parent education and suspension.
   
   b. Construct a crosstab for the variables “lowparenteduc” and “v7234” by selecting “analyze->descriptives->crosstabs.” Make sure to put the independent variable in the column box and to select “column cell percentages.”
   
   c. What does this tell us about the relationship between the two variables?
   
   d. Do we have sufficient evidence to say that this is a causal relationship? Which criteria of causality are/are not met?
   
   e. What else do you think we might want to account for to ensure that this is not a spurious relationship?

3. Let’s elaborate on this by accounting for a potential confounder. Some have argued that race is critical to understanding school discipline. They suggest that low parent education and suspension are both correlated with a parent/respondent’s race. Therefore, race may be a confounder in this relationship, raising the possibility that education doesn’t matter after all. We can control for the effect of race by constructing a layered crosstabs, which will look at the relationship between parent education and school suspension in whites and blacks separately.

   a. Go back to the crosstab menu. This time, add the variable “white” to the “layer” field.
   
   b. The result here will be a bit daunting at first glance, but what you are looking for is if the pattern you found earlier is present in the nonwhite subgroup and the white subgroup. If the effect is not present for one group, then you have a partial confounder, where the relationship between the two variables is only present for certain groups. If the relationship goes away when you account
for this new variable, then you have found a confounder and the original relationship might have been spurious.

c. In this case, is there no confounding, partial confounding, or total confounding?

d. Does having parents with low education influence the risk of suspension for whites? For blacks? What does this tell us about our society and how punishment is distributed in schools?

4. Based on the measures in this dataset, what type of design do you think this study is using: longitudinal or cross-sectional? What are the strengths and limitations of this approach?