Conducting a Good Experiment I: Variables and Control

CHAPTER

6

In this chapter we begin to examine the methods and procedures that will allow us to make cause-and-effect inferences. We begin by carefully examining the types of variables used by the researcher: independent, extraneous, dependent, and nuisance variables. Recall from Chapter 5 that the experimenter directly manipulates *independent* variables; *dependent* variables change in response to the independent variable manipulation; and *extraneous* variables can invalidate our experimental results. As we will see, *nuisance* variables cause our results to be less clear. Once we have discussed these variables and made their relation to the experiment clear, we will consider the procedures that researchers have developed to keep unwanted, extraneous variables from influencing the results of our experiment.

The Nature of Variables

Variable An event or behavior that can assume two or more values.

Before jumping into a discussion of independent variables (IV), let's look at the nature of variables in general. A **variable** is an event or behavior that can assume at least two values. For example, temperature is a variable; it can assume a wide range of values. The same could be said for height, weight, lighting conditions, the noise level in an urban area, anxiety, confidence, and your responses to a test, as well as many other possibilities; each of these events can assume two or more values or levels.

So, when we discuss variables involved in a psychological experiment, we are talking about events or behaviors that have assumed *at least two values*. If the IV has only one level, we would have nothing against which to compare its effectiveness. Assume you want to demonstrate that a new brand of toothpaste is the best on the market. You have a group of participants try the new toothpaste and then rate its effectiveness. Even though the entire group rates the toothpaste in question as "great," you still cannot claim that it is best; you do not have ratings from other groups using different brands.

Just as the IV must have at least two values, the dependent variable (DV) also must be able to assume two or more values. Your toothpaste study would be meaningless if the *only* response the participants can make is "great"; more than one response alternative is needed.

The same logic applies in the case of extraneous variables. If two or more values are not present, then the event in question is not an extraneous variable. If all the participants in our toothpaste study are women, then we do not have to be concerned with sex differences between the groups that we test. (This point will be important later in the Controlling Extraneous Variables section of this chapter.) Notice that our concern about extraneous variables is quite different from our concern about IVs and DVs. Whereas we were concerned that the IV and DV have or are able to assume two or more values, we seek to avoid those instances where extraneous variables can assume two or more values.

Operationally Defining Variables

As you will recall from Chapter 5, we suggested that replication of past research can be a valuable source of research ideas. Let's assume you have located a piece of research that you want to replicate. You carefully read how the experiment was conducted and find that each participant received a reward following every correct response. Assume this sentence is the only information you have concerning the *reward* and *response* involved in the experiment. If you asked 10 different researchers what reward they would use and what response they would record, how many different responses would you get? With this limited and vague information, chances are good that you would get as many different answers as the number of people you asked. How valid will your replication be? If you use a totally different reward and a totally different response, have you even conducted a replication?

Problems and concerns such as these led a 1920s Harvard University physicist, Percy W. Bridgman, to propose a way to obtain clearer communication among researchers and thus achieve greater standardization and uniformity in experimental methodology (Goodwin, 2005). Bridgman's suggestion was simple: Researchers should define their variables in terms of the operations needed to produce them (Bridgman, 1927). If you define your variables in this manner, then other scientists can replicate your research by following the definitions you have given for the variables involved; such definitions are called **operational definitions**. Operational definitions have been a cornerstone of psychological research for nearly three-quarters of a century because they allow researchers to communicate clearly and effectively with each other.

Operational definition Defining the independent, dependent, and extraneous variables in terms of the operations needed to produce them.

To illustrate the use of operational definitions, let's return to the reward and response situation we described previously. If we define reward as "a 45-mg Noyes Formula A food pellet," then other animal researchers can use the same reinforcer by ordering a supply of 45-mg Formula A pellets from the P J. Noyes Company. Likewise, if we define the response as "making a bar press in an operant conditioning chamber (Lafayette Model 81335)," then another researcher can replicate our research setup by purchasing a similar piece of equipment from the Lafayette Instrument Company.

The experimenter must be able to convey clearly such information about all the variables involved in a research project. Hence, it is crucial to give operational definitions for the IV, DVs, and extraneous variables, as well as for nuisance variables.

Independent Variables

Independent variables (IVs) are those variables that the experimenter purposely manipulates. The IV constitutes the reason the research is being conducted; the experimenter is interested in determining what effect the IV has. The term *independent* is used because the IV does not depend on other variables; it stands alone. A few examples of IVs that experimenters have used in psychological research are sleep deprivation, temperature, noise level, drug type (or dosage level), removal of a portion of the brain, and psychological context. Rather than attempting to list all possible IVs, it is easier to indicate that they tend to cluster in several general categories.

Types of IVs

Physiological IV A physiological state of the participant manipulated by the experimenter.

Physiological When the participants in an experiment are subjected to conditions that alter or change their normal biological state, a **physiological IV** is being used. For example, Susan Nash (1983), a student at Emporia State University in Emporia, Kansas, obtained several pregnant rats from an animal supplier. Upon their arrival at the laboratory, she randomly assigned half the rats to receive an alcohol–water mixture during gestation; the remainder received plain tap water. She switched the alcohol-exposed mothers to plain tap water when the pups were born. Thus, some rat pups were exposed to alcohol during gestation, whereas others were not. Nash tested all the pups for alcohol preference when they were adults and found that those animals that were exposed to alcohol (the physiological IV) during gestation drank more alcohol as adults. Nash received the 1983 J. P. Guilford–Psi Chi National Undergraduate Research Award for this experiment. Just as alcohol exposure was a physiological IV in Nash's experiment, administering a new drug to determine whether it is successful in alleviating schizophrenic symptoms also represents a physiological IV.

Experience IV Manipulation of the amount or type of training or learning.

Experience When the effects of amount or type of previous training or learning are the central focus of the research, the researcher is using an **experience IV**. A study conducted by Monica Boice, a student at Saint Joseph's University in Philadelphia, and her faculty advisor, Gary Gargano, illustrates the use of experience as an IV. Boice and Gargano (2001) studied memory for items in a list as a function of the number of related cues that were presented at the time of recall. Some participants received zero cues, whereas other participants received eight cues. The number of cues was an experience IV. The results of this study indicated that, under some conditions, receiving eight related cues actually resulted in worse memory performance than did receiving no cues.

Stimulus or **environmental IV** An aspect of the environment manipulated by the experimenter.

Stimulus Some IVs fall into the category of **stimulus** or **environmental variables**. When researchers use this type of IV, they are manipulating some aspect of the environment. Kathy Walter, Sammi Ervin, and Nicole Williamson, students at Catawba College in Salisbury, North Carolina, conducted a study under the direction of Sheila Brownlow in which they used a stimulus IV (Walter, Brownlow, Ervin, & Williamson, 1998). They asked 144 college students to judge various traits of women who walked barefoot and then wore high heels. The stimulus variable was whether the person being judged was barefoot or wore heels. The results showed that when the women wore heels, student participants judged them as less sexy and more submissive than when they were barefoot.

Participant It is common to find **participant characteristics**, such as age, sex, personality traits, or academic major, being treated *as if* they are IVs.

Participant characteristics Aspects of the participant, such as age, sex, or personality traits, which are treated as if they were IVs.

Although many researchers may treat participant characteristics as if they are IVs, they really are not. Why not?

To be considered an IV, the behavior or event in question must be directly manipulated by the experimenter. Although experimenters can manipulate physiological, experience, and stimulus IVs, they are not able to manipulate participant characteristics directly. For this reason, experimenters do not consider them to be true IVs. The experimenter does not create the participants′ sex or cause participants to be a certain age. Participant characteristics or variables are best viewed as *classification*, not manipulation, variables. The categories for participant variables are created before the experiment is conducted, and the experimenter simply assigns the participants to these categories on the basis of the characteristics they display.

Extraneous Variables (Confounders)

Extraneous variables are those factors that can have an *unintended* influence on the results of our experiment. Extraneous variables influence the difference *between* groups. Figure 6-1A shows the relation of two groups without the influence of an extraneous variable; two possible effects of an extraneous variable are shown in Figures 6-1B and 6-1C. Thus, an extraneous variable can unintentionally cause groups to move closer together (Figure 6-1B) or farther apart (Figure 6-1C).

Review Figure 6-1 and the information we have presented about extraneous variables. The effect of an extraneous variable is similar to that of another major component of the experiment. What role does an extraneous variable appear to play in an experiment? How is the presence of an extraneous variable detrimental to the experiment?

The other component of an experiment that can influence the difference between groups is the IV. Thus, an extraneous variable can affect the outcome of an experiment. Just as other likely interpretations can damage a detective's case beyond repair, the presence of an extraneous variable is devastating to research; it is not possible to attribute the results of the

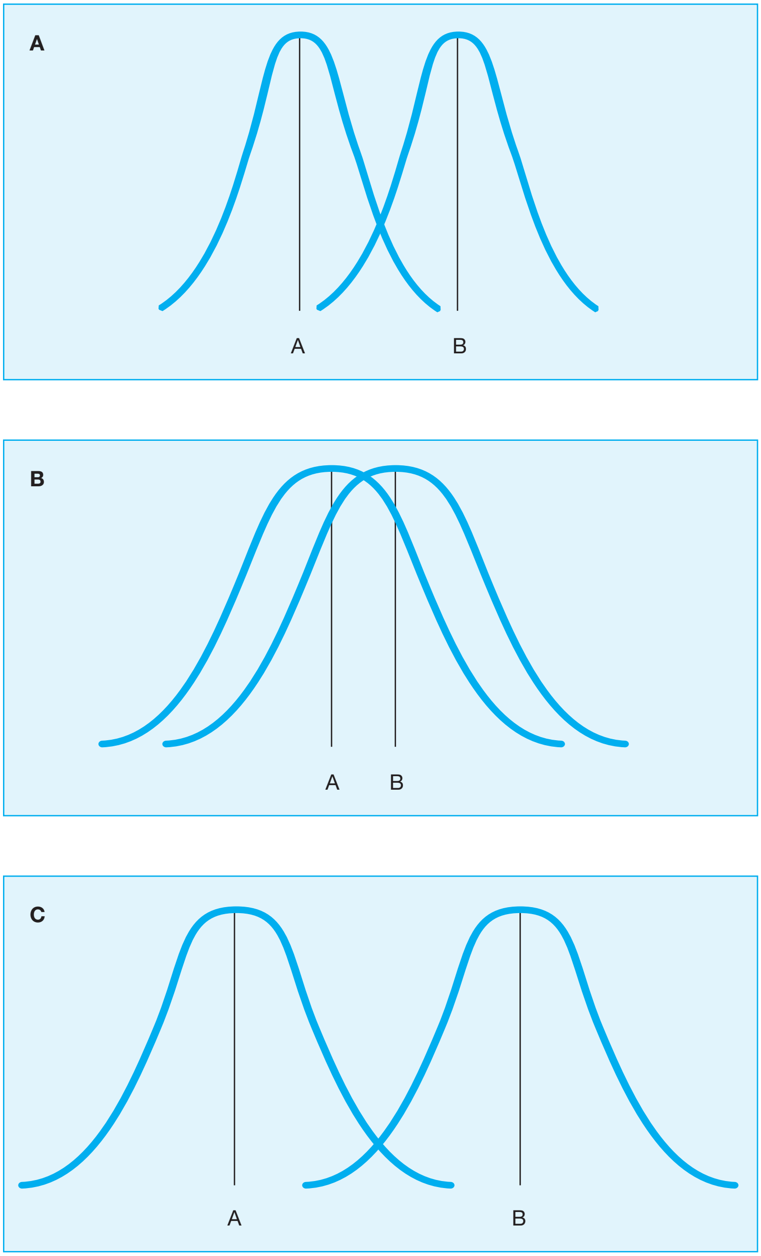

Figure 6-1 **A.** The difference (A = standard method; B = new method) between two groups with no confounder operating. **B.** The difference between two groups when a confounder is present and has moved the groups closer together. **C.** The difference between two groups when a confounder is present and has moved the groups farther apart.

experiment to the IV. Why? There are two variables that may have caused the groups to differ: the IV you manipulated *and* the unwanted extraneous variable. You have no way to determine which of these two variables caused the differences you observed. In such instances, when we can attribute the results either to an IV or to an extraneous variable, the experiment is **confounded**. (The terms *extraneous variable* and *confounder* are often used synonymously.) When an experiment is confounded, the best course of action is to discontinue the research and learn from your mistake. You can control the extraneous variable in the next experiment.

Confounding A situation in which the results of an experiment can be attributed to either the operation of an IV or an extraneous variable.

To illustrate how confounding works, let's consider a reading comprehension study. We have first- and second-graders available to serve as participants. The researcher assigns all the first-graders to the standard method for teaching reading comprehension and all the second-graders to the new method. The experimenter conducts the experiment and finds that the comprehension scores for students using the new method are substantially better than those of students using the standard method (see Figure 6-1C). What would have happened if the researcher had assigned the second-graders to the standard method and the first-graders to the new method? We might have seen results like those shown in Figure 6-1B. Why did these two sets of results occur? In each instance it is arguable that a preexisting difference in reading comprehension between the groups of children created differences between the two groups (i.e., the preexisting difference acted *as if* it were an IV). Assuming that second-graders have superior reading comprehension, it seems reasonable to suggest that the new method seemed even more effective when the second-graders used it (i.e., the difference between the groups was exaggerated). However, when the second-graders used the standard method, the superior method used by the first-graders increased their scores and moved the groups closer together (i.e., group differences decreased). Certainly, all of this commentary is only speculation on our part. It is also possible that the IV created the group differences that we observed. The main point is that we really do not know what caused the differences—the IV (type of method used) or the extraneous variable (grade level).

The presence of an extraneous variable is often very difficult to spot; it may take several knowledgeable individuals scrutinizing an experiment from every possible angle to determine whether one is present. If the experimenter detects an extraneous variable before conducting the research, then the experimenter can deal with the problem and proceed with the experiment. We present techniques for controlling unwanted variables later in this chapter.

Dependent Variables

The dependent variable (DV) changes as a function of the level of the IV experienced by the participant; therefore, the value the DV assumes *truly depends* on the IV. The DV consists of the data or results of our experiment. As with all aspects of psychological research, experimenters must give the DV appropriate consideration when they formulate an experiment. The experimenter must deal with such considerations as selecting the appropriate DV, deciding exactly which measure of the DV to use, and whether to record more than one DV.

Selecting the DV

Because psychology often is defined as the science of behavior, the DV typically consists of some type of behavior or response. When the researcher administers the IV, however, it is likely that several responses will occur. Which one should the researcher select as the DV? One answer to this question is to look carefully at the experimental hypothesis. If you have stated your hypothesis in general implication form ("if . . . then"—see Chapter 5), the "then" portion of the hypothesis will give you an idea of the general nature of your DV. For the Boice and Gargano (2001) memory study the choice was easy: "The dependent measure was the number of words correctly recalled" (p. 119).

What if your hypothesis were more general? Say that you wanted to study "spatial abilities." Where could you find information to help you choose a *specific* DV? We hope you are already a step ahead of us; our literature review (see Chapter 2) can provide valuable guidelines. If other researchers have used a particular response successfully as a DV in previous research, chances are that it will be a good choice again. Another reason for using a DV that has been used previously by researchers is that you will have a comparison for your results. Although totally different DVs may provide exciting new information, the ability to relate the results of experiments using different responses is more difficult.

Recording or Measuring the DV

After you have selected the DV, you will have to decide exactly how to measure or record it. Several possibilities exist.

Correctness With this DV measure, the participant's response is either correct or incorrect. Because they counted the number of words their participants remembered, Boice and Gargano (2001) used a correctness DV.

Rate or Frequency If you were studying the lever-pressing performance of a rat or pigeon in an operant conditioning chamber (Skinner box), then your DV would likely be the rate of responding shown by the animal. The rate of responding determines how rapidly responses are made during a specified time period. You can plot your data in the form of a cumulative record with steeper slopes representing higher rates (i.e., large numbers of responses being made in shorter periods of time). Figure 6-2 shows some different rates of responding.

If you were studying the number of social interactions among children during free play at a kindergarten, you might want to record the frequency, rather than the rate, of responding. Your DV, then, would simply be the number of responses shown during a specified time period without any concern for how rapidly the participant makes them.

Degree or Amount Often researchers record the DV in terms of degree or amount. In this instance, you do not record the number or frequency of the participant's responses; rather, you typically record a single number that indicates the degree or amount. Amie McKibban and Shawn Nelson, students at Emporia State University in Emporia, Kansas, studied satisfaction with life in college students (McKibban & Nelson, 2001). Scores on a Satisfaction With Life scale measured how satisfied (i.e., degree or amount) their participants were with life.

Latency or Duration In many situations, such as studies of learning and memory, how quickly participants make a response (latency) or how long the response lasts (duration) are of particular interest. For example, Rachel Ball, Erica Kargl, J. Davis Kimpel, and Shana Siewert, students at Wheaton College in Wheaton, Illinois, were interested in the relation between a participant's mood and his or her reaction time measured as a latency DV. They found that

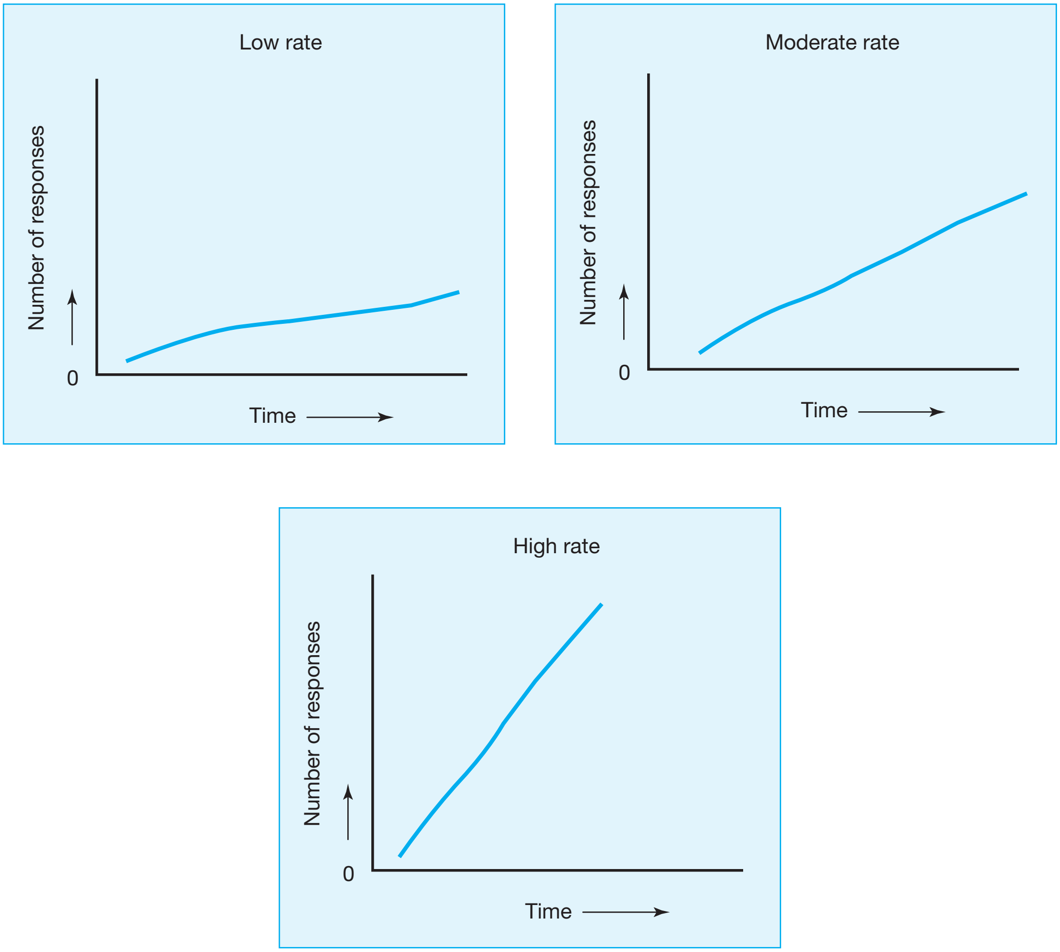

Figure 6-2 Different Rates of Responding.

participants in sad and suspenseful moods had longer reaction times than participants in a neutral mood (Ball, Kargl, Kimpel, & Siewert, 2001).

Recording More Than One DV

If you have the measurement capabilities, there is nothing to prohibit the recording of more than one DV. Possibly additional data will strengthen your knowledge claim in the same way it might strengthen a detective's case. Should you record additional DVs? The answer to this question really boils down to deciding whether recording additional DVs is going to add appreciably to your understanding of the phenomenon under study. If recording an additional DV makes a meaningful contribution, then you should give it serious consideration. If measuring and recording another DV does not make a substantive contribution, then it is probably not worth the added time and trouble. Often you can use previous research as a guide concerning whether you should consider recording more than one DV.

PSYCHO-LOGICAL DETECTIVE

Consider a reversed eye-hand coordination (mirror tracing) experiment in which you record the time taken to complete the tracing of a star pattern (a latency DV) while looking in a mirror. Is this DV sufficient to give you a good, complete picture of the performance of this task, or should you also record another DV?

You probably should record a second DV. The latency DV indicates only how long it took to trace the star pattern. The experimenter has no record of the number of errors (going beyond the boundaries of the figure) made by the participants. A second DV, which measures the number of errors (a frequency DV), will make a significant contribution to this experiment.

The need for more than one DV was recognized in an experiment conducted by Janet Luehring, a student at Washburn University in Topeka, Kansas, and her faculty advisor, Joanne Altman (Luehring & Altman, 2000). These investigators studied male–female differences in spatial ability performance. Their participants performed a mental rotation (visualizing what an object would look like after it was rotated in space). Because the participant's performance could be correct, incorrect, or (because the task was timed) uncompleted, Luehring and Altman recorded *both correct and incorrect responses*. The number of incorrect responses, which was *not* equal to the total number of responses minus the numer of correct responses, had the potential to provide additional, relevant information.

Characteristics of a Good DV

Although considerable thought may go into deciding exactly how the DV will be measured and recorded and whether more than one DV should be recorded, the experimenter still has no guarantee that a good DV has been selected. What constitutes a good DV? We want the DV to be valid.

Valid Measuring what is supposed to be measured.

The DV is **valid** when it measures what the experimental hypothesis says it should measure. For example, assume you are interested in studying intelligence as a function of the differences in regional diet. You believe the basic diet consumed by people living in different regions of the United States results in differences in intelligence. You devise a new intelligence test and set off to test your hypothesis. As the results of your project start to take shape, you notice that the scores from the Northeast are higher than those from other sections of the country; your hypothesis appears to be supported. However, a closer inspection of the results indicates that participants not living in the Northeast miss only certain questions. Are all the questions fair and unbiased, or do some favor Northeasterners? For example, you notice there are several questions about subways. How many individuals from Arizona are familiar with subways? Thus, your DV (scores on the intelligence test) may have a regional bias and may not measure the participants' intelligence consistently from region to region. A good DV must be directly related to the IV and must measure the effects of the IV manipulation as the experimental hypothesis predicts it will.

Reliable Producing consistent measurements.

A good DV is also **reliable**. If the scores on an intelligence test are used as the DV, then we would expect to see similar scores when the test is administered again under the same IV conditions (test–retest procedure; see Chapter 4). If the test gives the same individual different IQ scores at different times, it is not a reliable test.

Nuisance Variables

Nuisance variable Unwanted variables that can cause the variability of scores within groups to increase.

Nuisance variables are either characteristics of the participants or unintended influences of the experimental situation that make the effects of the IV more difficult to see or determine. It is important to understand that nuisance variables influence all groups in an experiment; their influence is not limited to one specific group. When they are present, nuisance variables result in greater variability in the DV; the scores *within* each group spread out more. For example, assume you are interested in studying reading comprehension. Can you think of a participant characteristic that might be related to reading comprehension? How about intelligence or IQ?

Figure 6-3A shows the spread of the reading comprehension scores within a group when there are not wide differences in intelligence among the participants within the group. In this instance a nuisance variable is not operating.

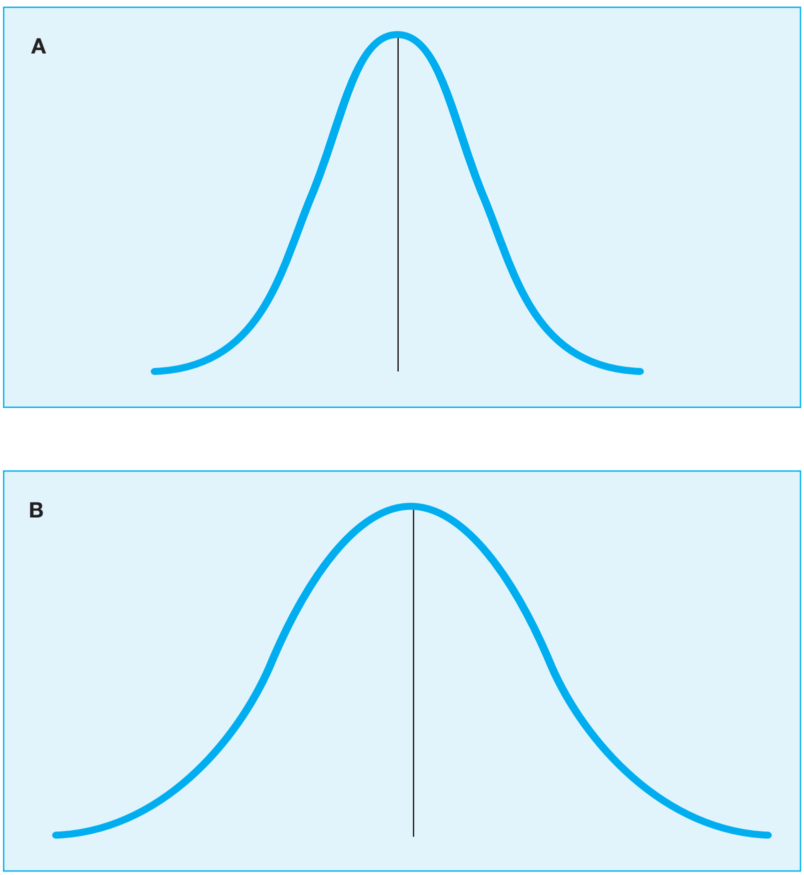

Figure 6-3 **A.** The spread of scores within a group when a nuisance variable is not operating. **B.** The spread of scores within a group when a nuisance variable is operating.

You can see how the majority of the comprehension scores are similar and cluster in the middle of the distribution; there are relatively few extremely low and extremely high scores. Figure 6-3B shows the distribution of comprehension scores when there are wider differences in intelligence (i.e., a nuisance variable is present). Notice that the scores are more spread out; there are fewer scores in the middle of the distribution and more scores in the extremes.

How does a nuisance variable influence the results of an experiment? To answer that question, we need to add another group of participants to our example and conduct a simple experiment. Imagine we are evaluating two methods for teaching reading comprehension: the standard method and a new method. In Figure 6-4A, we are comparing two groups that have not been influenced by the nuisance variable. The difference between these two groups is pronounced and clear; they overlap very little.

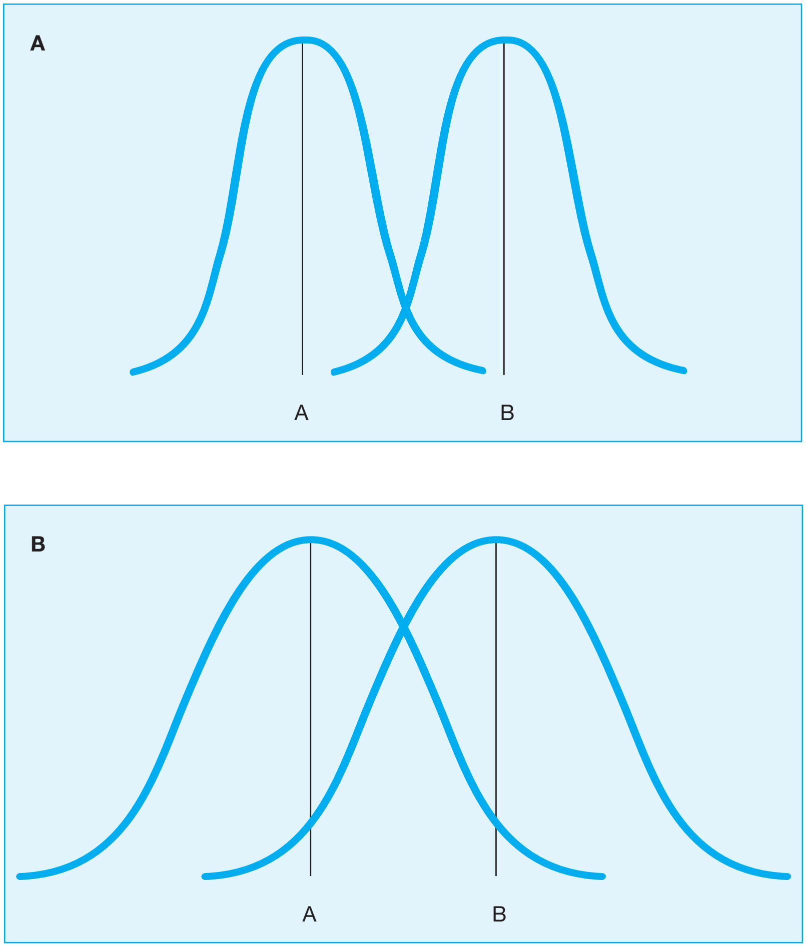

FIGURE 6-4 **A.** A Comparison of Two Groups When a Nuisance Variable Is Not Operating. **B.** A Comparison of the Same Two Groups When a Nuisance Variable Is Operating.

Let's add the effects of a nuisance variable, such as wide differences in verbal ability, to each group and then compare the groups. As you can see in Figure 6-4B, when the scores spread out more, there is greater overlap, and the difference between the groups is not as distinct and clear as when the nuisance variable was not present. When a nuisance variable is present, our view of the experimental results is clouded; we are unable to see clearly the difference the IV may have created between the groups in our experiment. Notice that when the nuisance variable was added (Figure 6-4B), the *only* thing that happened was that the scores spread out in both extremes of each distribution—the relative location of the distributions did not change. *Nuisance variables increase the spread of scores within a distribution; they do not cause a distribution to change its location*.

For each example, indicate the nuisance variable and its effect.

1. An experimenter measures reaction time in participants ranging in age from 12 to 78 years.
2. The ability of participants to recall a list of words is being studied in a room that is located by a noisy elevator.
3. The laboratory where participants are tested for manual dexterity has frequent, unpredictable changes in temperature.

In the first situation the wide age range is the nuisance variable. The younger participants should display faster reaction times than the older participants; therefore, the scores will spread into both ends of the distributions. The change in noise level caused by the operation of the elevator is the nuisance variable in the second example; the frequent, unpredictable temperature changes are the nuisance variable in the third example. The change in conditions in all three examples is likely to increase the spread of scores. Our goal as researchers is to keep nuisance variables to a minimum so that the effects of the IV are as clear as possible.

REVIEW SUMMARY

1. A **variable** is an event or behavior that can assume at least two levels.
2. **Independent variables** are purposely manipulated by the experimenter and form the core or purpose of the experiment. **Physiological IVs** refer to changes in the biological state of the participants, whereas **experience IVs** refer to manipulations of previous experience or learning. **Stimulus IVs** are manipulations in some aspect of the environment.
3. Although **participant characteristics** such as age, sex, or personality traits are often treated as IVs, technically they are not IVs because the experimenter does not directly manipulate them.
4. **Extraneous variables** (**confounders**) can have an unintended influence on the results of an experiment by changing the difference *between* the groups. When an extraneous variable is present, the experiment is confounded.
5. The **dependent variable** changes as a function of the changes in the IV. The experimental hypothesis can provide possible guidelines concerning the selection of the DV. Past research also can assist the experimenter in selecting the DV.

6. The DV can be recorded in terms of correctness, rate or frequency, degree or amount, and latency or duration. If additional information will be gained, the experimenter should consider recording more than one DV. A good DV is directly related to the IV (**valid**) and **reliable**.
7. **Nuisance variables** are variables that increase the variability of scores *within all groups*. The presence of nuisance variables makes the results of an experiment less clear.

Check Your Progress

1. An event or behavior that can assume at least two values is a ________.
2. Matching

1. DV	A. change in normal biological state
2. extraneous variable	B. manipulation of environment
3. physiological IV	C. can damage the experiment and its results
4. experience IV	D. age
5. stimulus IV	E. changes as a function of changes in IV
6. participant variable	F. amount of previous learning

3. Your research involves determining the effects of persuasion on the strength of attitudes. You are using a ________ measurement of the DV.
   a. correctness
   b. rate
   c. degree
   d. duration
4. A good DV has two primary qualities; it is both ________ and ________.
   a. easy to identify; easy to measure
   b. dependent on the experimenter; identifiable
   c. valid; reliable
   d. positively correlated; negatively correlated
5. Under what conditions should you record more than one DV?
6. Variables that result in greater within-group variability in the data are called
   a. independent variables
   b. confounders
   c. nuisance variables
   d. uncontrolled variables

Controlling Extraneous Variables

Just as care and precision are crucial when a detective attempts to solve a case, control forms an integral component of psychological research. The experimenter must exercise control over both extraneous variables and nuisance variables so that the results of the experiment are as meaningful (no extraneous variables present) and clear (minimal influence of nuisance variables) as possible. When you are dealing with a variable that can be clearly specified and quantified (e.g., sex, age, educational level, temperature, lighting intensity, or noise level),

"It was more of a 'triple-blind' test. The patients didn't know which ones were getting the real drug, the doctors didn't know, and I'm afraid, nobody knew."

Although psychologists can use many different types of experimental control in their research, none of them should be this complicated.

one of the five basic control techniques should be applicable. We describe these five basic control techniques—randomization, elimination, constancy, balancing, and counterbalancing—in the next section.

Basic Control Techniques

As we discuss the basic control techniques, it is important to keep in mind that their goals are to (a) produce groups that are equivalent prior to the introduction of the IV, thereby eliminating extraneous variables, and (b) reduce the effects of nuisance variables as much as possible.

Randomization We begin our discussion of control with randomization because it is the most widely used technique. **Randomization** guarantees that each participant has an equal chance of being assigned to any group in the experiment. For example, once the students had volunteered for the research on memory, Boice and Gargano (2001) indicated that "[t]he participants were placed randomly into one of the six experimental groups" (p. 120).

Randomization A control technique that ensures that each participant has an equal chance of being assigned to any group in an experiment.

The logic behind using randomization is as follows. Because all participants have an equal likelihood of being selected for each group in an experiment, any unique characteristics associated with the participants should be equally distributed across all groups that are formed. Consider level of motivation, for example. Although it may not be feasible to measure each participant's level of motivation, this variable can still be controlled by randomly forming the groups in our experiment. Just by chance we would expect that *each* group would have some participants who are highly motivated, some participants who are moderately motivated, and some participants who are barely motivated. Thus, the groups should be equated with regard to the average level of motivation, as well as the myriad of other unknown and unsuspected extraneous variables that might be present.

PSYCHO-LOGICAL DETECTIVE

Even though it is the most widely used control procedure, randomization has one major drawback. What is it?

Because we are never *fully* aware of the variables that randomization controls and thus do not know whether these variables are distributed evenly to all groups, we cannot be positive that this control technique is effective. It is possible that all the highly motivated participants will be randomly assigned to the same group. Also, you need to consider what is being controlled when randomization is used. If you find yourself having some difficulty specifying exactly what is controlled by randomization, then you are on the right track. Randomization is used as a control technique for *all* variables that *might* be present and of which the experimenter is unaware. If the experimenter is unaware of exactly which variables are being controlled and how effective the control is, then it should come as no surprise that the experimenter is never *completely* sure how effective randomization has been.

Elimination A control technique whereby extraneous variables are completely removed from an experiment.

Elimination When we know the extraneous variables or nuisance variables, our approach can be more direct. For example, we might choose to remove or **eliminate** the unwanted variable. This sounds easy, but in practice you may find it quite difficult to remove a variable *completely*.

Shann Sagles, Sharon Coley, Germilina Espiritu, and Patricia Zahregian, students at Pace University in White Plains, New York, and their faculty advisor, Richard Velayo, used elimination as a control in their cross-cultural study of the identification of facial expressions. "The 35-mm photos of the target individuals, taken from the base of the chin to the top of the forehead, excluded their attire and body type. The rationale for this procedure was to *eliminate* [emphasis added] extraneous variables that may have influenced the participants' responses" (Sagles, Coley, Espiritu, Zahregian, & Velayo, 2002, p. 33). Thus, Sagles et al. guaranteed that clothing and weight did not affect their participants' responses.

When the variable in question consists of an entire category of events, such as noise, temperature, or lighting condition, it may be difficult, if not impossible, to eliminate it. If, however, the variable is a specific occurrence within one of these more general categories, such as temperatures above 80 degrees, then it may be possible to eliminate that *aspect* of the variable. In this situation, however, the experimenter is interested not just in eliminating a variable but also in producing and maintaining a *constant* condition under which the participants in the experiment are tested.

Constancy A control technique by which an extraneous variable is reduced to a single value that is experienced by all participants.

Constancy When it is difficult or impossible to eliminate a variable completely, the experimenter may choose to exercise control by creating a uniform or constant condition which is experienced by all participants. **Constancy** has become a standard control technique for many researchers. For example, experimental testing may take place in the same room, with the same lighting and temperature levels, and at the same time of day (if the experiment is conducted on more than one day). In this instance the location of the experiment, the temperature level, the lighting level, and the time of day have not been eliminated but rather have assumed a constant value.

Juan and Nancy are interested in determining which of two methods of teaching psychological statistics is best. Two statistics classes are available: Juan teaches one method to one class, and Nancy teaches the second method to the other class. This experiment is confounded because Juan and Nancy each teach one of the classes, so it is impossible to tell whether differences between the classes are due to the method or the teacher. How would you use constancy to control this extraneous variable?

The easiest approach would be to have only one of the experimenters, Juan or Nancy, teach both classes. Thus, the extraneous variable, the teacher, is the same for both classes and the experiment is no longer confounded.

By making sure that the experimental testing conditions do not vary unpredictably, constancy can also control for nuisance variables. When the testing conditions are the same from testing session to testing session, there is a greater likelihood that the scores within the groups will not spread out as much because of the constancy. Constancy can also control nuisance variable effects produced by participant variables such as age, sex, and educational level. For example, in a study of the effects of massage and touch on body dissatisfaction, Angela Larey, a student at Missouri Southern State College in Joplin, Missouri, "used only women [as participants] because they generally manifest greater body dissatisfaction than men" (Larey, 2001, p. 79). Recall that wide variations in a such participant variables may result in a greater spread of scores within the groups. If the technique of stratified random sampling (see Chapter 5) is used, then the variability among the participants' scores should be smaller because the participants are more homogeneous. Clearly, this procedure helps create constancy.

Although constancy can be an effective control technique, there are situations in which the unwanted variable(s) cannot be reduced to a single value that all participants experienced in the experiment. What can be done when this variable assumes two or more values? The answer may lie in the control technique known as balancing.

Balancing **Balancing** represents a logical extension of control through constancy. Thus, the groups in our experiment are balanced or equivalent when each group experiences *all* unwanted variables or *levels* of unwanted variables in the same manner or to the same degree.

Balancing A control procedure that achieves group equality by distributing extraneous variables equally to all groups.

In the simplest example of balancing we would test two groups—one group (the experimental group) would receive the IV; the second group (the control or comparison group) would be treated identically but would not receive the IV. If the groups were balanced or equated with regard to extraneous variables, then we could tentatively conclude that differences between them were caused by the IV. This general situation is diagrammed in Table 6-1.

When the potential extraneous variables, such as various personality differences in human participants, are *unknown*, the experimenter uses randomization to form equivalent groups, and we assume that the respective extraneous variables are distributed equally to all groups. When the extraneous variables, such as sex of the experimenter, are known, then the experimenter can be more systematic in the use of the balancing technique to produce equivalent conditions.

TABLE 6-1 Balanced Extraneous Variables

When the extraneous variables are experienced in the same manner, the groups are said to be balanced.

Group 1	Group 2
Treatment A	Treatment B
Ext. Var. 1	Ext. Var. 1
Ext. Var. 2	Ext. Var. 2
Ext. Var. 3	Ext. Var. 3
Ext. Var. 4	Ext. Var. 4

Let's assume that all the students included in Juan and Nancy's experiment cannot be taught by one teacher; two are needed. How could balancing be used to remove the potential confounding caused by a different teacher teaching each of the new methods?

As you can see in Table 6-2, an easy solution to their problem is to have Juan and Nancy each teach half the students under *each* method. Thus, the two teachers appear equally under each teaching method and the classes are balanced with regard to that potential confounding variable. This teaching example illustrates the simplest situation in which balancing is used to control one extraneous variable. Balancing can also be used with several extraneous variables (see Table 6-1); the only requirement is that each extraneous variable appear equally in each group.

Although elimination, constancy, and balancing offer the experimenter powerful control techniques, they are not able to deal with all control problems. In the next section we examine one of these problem situations, *sequencing* or *order*, and how to control it through counterbalancing.

TABLE 6-2 Using Balancing to Eliminate Confounding in Teaching Two Methods of Psychological Statistics

Juan	Nancy
25 students → Method 1	25 students → Method 1
25 students → Method 2	25 students → Method 2

Counterbalancing In some experiments, participants participate in more than one experimental condition. For example, you might want to conduct a cola taste test to determine which of two brands of cola is preferred. As you set up your tasting booth at the local mall, you are sure you have taken all the right precautions: The tasting cups are all the same, the two colas will be poured from identical containers, the participants will consume the same amount of Cola A and then Cola B (in that order), and the participants will be blindfolded during the test so color differences will not influence their choice. Your control seems to be perfect. Is it? *Constancy* is achieved by ensuring that (a) the tasting cups are the same, (b) the colas are poured from similar containers, and (c) all participants consume the same amount of each cola. By blindfolding the participants, you have *eliminated* any problems that may be caused by differences in the visual appearance of the two colas. These are all relevant control procedures.

PSYCHOLOGICAL DETECTIVE

Review the experiment we have just described. What control procedures are being used? A problem that has been overlooked needs to be controlled. What is this problem and how might it be controlled?

The problem that is overlooked concerns the sequence or order for sampling the two colas. If Cola A is *always* sampled before Cola B and one cola is liked more, you cannot be sure the preference is due to its great flavor *or* the fact that the colas were always sampled in the same order and this order may have influenced the participants' reactions. The technique used when a sequence or order effect must be controlled is known as **counterbalancing**. There are two types of counterbalancing: within-subject and within-group. **Within-subject counterbalancing** attempts to control the sequence effect for each participant, whereas **within-group counterbalancing** attempts to control this problem by presenting different sequences to different participants.

Counterbalancing A procedure for controlling order effects by presenting different treatment sequences.

Within-subject counterbalancing Presentation of different treatment sequences to the same participant.

Within-group counterbalancing Presentation of different treatment sequences to different participants.

Within-Subject Counterbalancing Returning to the problem of sequence in our cola challenge, we could deal with this problem by having each participant sample the two colas in the following sequence: ABBA. By using within-subject counterbalancing, each participant will taste Cola A once before and once after tasting Cola B. Thus, the experience of having tasted Cola A first is counterbalanced by tasting Cola A last.

Although it may seem relatively easy to implement within-subject counterbalancing, there is one major drawback to its use: Each participant must experience each condition more than once. In some situations the experimenter may not want or be able to present the treatments more than once to each participant. For example, you may not have sufficient time to conduct your cola challenge and allow each participant the opportunity to sample *each* brand of cola more than once. In such instances within-group counterbalancing may offer a better control alternative.

Within-Group Counterbalancing Another way to deal with the cola challenge sequencing problem is randomly to assign half the participants to experience the two colas in the Cola A → Cola B sequence and the remaining half of the participants to receive the Cola B → Cola A sequence. The preference of participants who tasted Cola A before Cola B could be compared with the preference of the participants who tasted Cola B first.

Assuming that we tested six participants, the within-group counterbalanced presentation of the two colas would be diagrammed as shown in Table 6-3.

Table 6-3 Within-Group Counterbalancing for the Two-Cola Challenge When Six Participants Are Tested

	Tasting 1	Tasting 2
Participant 1	A	B
Participant 2	A	B
Participant 3	A	B
Participant 4	B	A
Participant 5	B	A
Participant 6	B	A

As you can see, three participants receive the A → B sequence, whereas three participants receive the B → A sequence. This basic diagram illustrates the three requirements of within-subject counterbalancing:

1. Each treatment must be presented to each participant an equal number of times. In this example, each participant tastes Cola A once and Cola B once.
2. Each treatment must occur an equal number of times at each testing or practice session. Cola A is sampled three times at Tasting 1 and three times at Tasting 2.
3. Each treatment must precede and follow each of the other treatments an equal number of times.

In this example, Cola A is tasted first three times and is tasted second three times.

Counterbalancing is not limited to two-treatment sequences. For example, let's assume that your cola challenge involves three colas instead of two. The within-group counterbalancing needed in that situation is diagrammed in Table 6-4. Carefully examine this diagram. Does it satisfy the requirements for counterbalancing?

It appears that all the requirements have been met. Each cola is tasted an equal number of times (6), is tasted an equal number of times (2) at each tasting session, and precedes and follows

Table 6-4 Within-Group Counterbalancing for the Three-Cola Challenge When Six Participants Are Tested

	Tasting Session		
	1	**2**	**3**
Participant 1	A	B	C
Participant 2	A	C	B
Participant 3	B	A	C
Participant 4	B	C	A
Participant 5	C	A	B
Participant 6	C	B	A

each of the other colas an equal number of times (2). It is important to note that, if we want to test more participants, they would have to be added in multiples of 6. The addition of any other number of participants violates the rules of counterbalancing by creating a situation in which one cola is tasted more than the others, appears more often at one tasting session than the others, and does not precede and follow the other colas an equal number of times. Try adding only 1 or 2 participants to Table 6-4 to see whether you can satisfy the requirements for counterbalancing.

Table 6-4 illustrates another consideration that must be taken into account when counterbalancing is used. When only a few treatments are used, the number of different sequences that will have to be administered remains relatively small and counterbalancing is manageable. When we tested 2 colas, only 2 sequences were involved (see Table 6-3); however, the addition of only 1 more cola resulted in the addition of 4 sequences (see Table 6-4). If we added an additional cola to our challenge (Colas A, B, C, and D), we would now have to administer a total of 24 different sequences, and our experiment would be much more complex to conduct. Our minimum number of participants would be 24, and if we wanted to test more than 1 participant per sequence, participants would have to be added in multiples of 24.

How do you know how many sequences will be required? Do you have to write down all the possible sequences to find out how many there are? No. You can calculate the total number of sequences by using the formula *n*! (*n* factorial). All that is required is to take the number of treatments (*n*), factor or break that number down into its component parts, and then multiply these factors or components. For example,

$$2! \text{ would be } 2 \times 1 = 2$$

$$3! \text{ would be } 3 \times 2 \times 1 = 6$$

$$4! \text{ would be } 4 \times 3 \times 2 \times 1 = 24$$

and so forth. When you can administer all possible sequences, you are using **complete counterbalancing**. Although complete counterbalancing offers the best control for sequence or order effects, often it cannot be attained when several treatments are included in the experiment. As we just saw, the use of 4 colas would require a minimum of 24 participants ($4! = 4 \times 3 \times 2 \times 1 = 24$) for complete counterbalancing. Testing 5 colas would increase the number of sequences (and the minimum number of participants) to 120 ($5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$). In situations requiring a large number of participants to implement complete counterbalancing, either you can reduce the number of treatments until your time, financial, and participant resources allow complete counterbalancing, or you can complete the experiment without completely counterbalancing.

Complete counterbalancing All possible treatment sequences are presented.

Incomplete counterbalancing refers to the use of some, but not all, of the possible sequences. Which sequences are to be used, and which ones are to be excluded? Some experimenters randomly select the sequences they will employ. As soon as the number of participants to be tested has been determined, the experimenter randomly selects an equal number of sequences. For example, Table 6-5 illustrates a possible random selection of sequences for conducting the cola challenge with four colas and only 12 participants.

Incomplete counterbalancing Only a portion of all possible sequences are presented.

Although random selection appears to be an easy approach to the use of incomplete counterbalancing, there is a problem. If you examine Table 6-5 carefully, you will see that although each participant receives each treatment an equal number of times, the other

TABLE 6-5 Incomplete Counterbalancing Using Randomly Selected Tasting Sequences for the Four-Cola Challenge Using 12 Participants

	Testing Session			
	1	2	3	4
Participant 1	A	B	C	D
Participant 2	A	B	D	C
Participant 3	A	C	D	B
Participant 4	A	D	C	B
Participant 5	B	A	C	D
Participant 6	B	C	D	A
Participant 7	B	C	A	D
Participant 8	C	A	B	D
Participant 9	C	B	A	D
Participant 10	C	D	B	A
Participant 11	D	B	C	A
Participant 12	D	C	B	A

requirements for counterbalancing are not satisfied. Each treatment does not appear an equal number of times at each testing or practice session, and each treatment does not precede and follow each of the other treatments an equal number of times.

Two approaches can be adopted to resolve this problem, although neither one is completely satisfactory. We could randomly determine the treatment sequence for the *first* participant and then systematically rotate the sequence for the remaining participants. This approach is diagrammed in Table 6-6.

Thus, the first participant would taste the colas in the order B, D, A, C, whereas the second participant would experience them in the order D, A, C, B. We would continue systematically rotating the sequence until each cola appears once in each row and each column. To test a

TABLE 6-6 An Incomplete Counterbalancing Approach

This approach involves randomly determining the sequence for the first participant and then systematically rotating the treatments for the following sequences.

	Tasting Sequence			
	1	2	3	4
Participant 1	B	D	A	C
Participant 2	D	A	C	B
Participant 3	A	C	B	D
Participant 4	C	B	D	A

total of 12 participants, we would assign 3 participants to each of the 4 sequences. By ensuring that each treatment appears an equal number of times at each testing session, this approach comes close to satisfying the conditions for counterbalancing. It does not, however, ensure that the treatments precede and follow each other an equal number of times. A more complex procedure, the Latin square technique, is used to address this issue. Because of its complexity, this procedure is seldom used. If you are interested in reading about its use, Rosenthal and Rosnow (1991) offer a nice presentation of its particulars.

Now that we have examined the mechanics involved in implementing complete and incomplete counterbalancing, let's see exactly what counterbalancing can and cannot control. To say simply that counterbalancing controls for sequence or order effects does not tell the entire story. Although counterbalancing controls for sequence or order effects, as you will see, it also controls for *carryover effects*. It cannot, however, control for *differential carryover*.

Sequence or Order Effects **Sequence or order effects** are produced by the participant's being exposed to the sequential presentation of the treatments. For example, assume we are testing reaction time to three types of dashboard warning lights: red (R), green (G), and flashing white (FW). As soon as the warning light comes on, the participant is to turn off the engine. To counterbalance this experiment completely would require 6 sequences and at least 6 participants ($3! = 3 \times 2 \times 1 = 6$). If we found the reaction time to the first warning light, regardless of type, was 10 seconds, and that increases in reaction time of 4 and 3 seconds were made to the second and third lights (regardless of type), respectively, we would be dealing with a sequence or order effect. This example is diagrammed in Table 6-7. As you can see, the sequence or order effect depends on *where* in

Sequence or order effects The position of a treatment in a series determines, in part, the participants' response.

TABLE 6-7 Example of Sequence or Order Effects in a Counterbalanced Experiment

The performance decrease shown in parentheses below each sequence indicates the effect of testing reaction time to red (R), green (G), and flashing white (FW) lights on an instrument panel at that particular point in the sequence. Thus, second and third testings result in increases (i.e., slower reaction times) of 4 and 3, respectively, regardless of the experimental task.

	Order of Task Presentation		
	R	G	FW
Performance Increase →	(0	4	3)
	R	FW	G
	(0	4	3)
	G	R	FW
	(0	4	3)
	G	FW	R
	(0	4	3)
	FW	R	G
	(0	4	3)
	FW	G	R
	(0	4	3)

the sequential presentation of treatments the participant's performance is evaluated, not *which* treatment is experienced.

Sequence or order effects will be experienced equally by all participants in counterbalanced situations because each treatment appears an equal number of times at each testing session. This consideration points to a major flaw in the use of randomized, incomplete counterbalancing: The treatments may not be presented an equal number of times at each testing session (see Table 6-5). Thus, sequence or order effects are not controlled in this situation.

Carryover effect The effects of one treatment persist or carry over and influence responses to the next treatment.

Carryover Effects When a **carryover effect** is present, the effects of one treatment continue to influence the participant's response to the next treatment. For example, let's assume that experiencing the green (G) warning light before the red (R) light always causes participants to *decrease* their reaction time by 2 seconds. Conversely, experiencing R before G causes participants to *increase* their reaction time by 2 seconds. Experiencing the flashing white (FW) warning light either before or after G has no effect on reaction time. However, experiencing R before FW increases reaction time by 3 seconds, and experiencing FW before R reduces reaction time by 3 seconds. In the R → G/G → R and R → FW/FW → R transitions, the previous treatment influences the participant's response to the subsequent treatment in a consistent and predictable manner. These effects are diagrammed in Table 6-8. Note that counterbalancing includes an

TABLE 6-8 Example of Carryover Effects in a Counterbalanced Experiment

Carryover effects occur when a specific preceding treatment influences the performance in a subsequent treatment. In this example, experiencing Treatment G prior to Treatment R results in a decrease of 2 (i.e., −2), whereas experiencing Treatment R prior to Treatment G results in an increase of 2 (i.e., +2). Experiencing Treatment G prior to FW or Treatment FW prior to G does not produce a unique effect. However, experiencing Treatment R prior to FW results in an increase of 3, whereas experiencing Treatment FW prior to R results in a decrease of 3.

	Sequence of Treatments		
	G	R	FW
Effect on Performance →	(0	−2	+3)
	G	FW	R
	(0	0	−3)
	R	G	FW
	(0	+2	0)
	R	FW	G
	(0	+3	0)
	FW	G	R
	(0	0	−2)
	FW	R	G
	(0	−3	+2)

equal number of each type of transition (e.g., R → G, G → R, R → FW, etc.). Thus the opposing carryover effects cancel each other.

Differential Carryover Although counterbalancing can control many things, it offers no protection against differential carryover. **Differential carryover** occurs when the response to one treatment depends on *which* treatment is experienced previously. Consider an experiment investigating the effects of reward magnitude on reading comprehension in second-grade children. Each child reads three similar passages. After each passage is completed, a series of questions is asked. Each correct answer is rewarded by a certain number of M&Ms (the IV). In the low-reward condition (A), children receive one M&M after each correct answer; three and five M&Ms are received in the medium-reward (B) and high-reward (C) treatments, respectively.

Differential carryover The response to one treatment depends on which treatment was administered previously.

Although this experiment might be viewed as another six-sequence example of counterbalancing (see Table 6-4), the effects may not be symmetrical as in the carryover example we just considered. The participant who receives the A → B → C sequence may be motivated to do well at all testing sessions because the reward progressively increases from session to session. What about the student who receives the B → C → A sequence? In this case the reward rises from three M&Ms (Session 1) to five M&Ms (Session 2), but then is reduced to one M&M (Session 3). Will the decrease from five M&Ms to one M&M produce a unique effect—such as the student's refusing to participate—not seen in the other transitions? If so, differential carryover has occurred, and counterbalancing is not an effective control procedure. Some possible effects of differential carryover in the M&M study are shown in Table 6-9. As you can see, the

TABLE 6-9 Example of Differential Carryover in a Counterbalanced Experiment

Differential carryover occurs when performance depends on which specific sequence occurs. In the following example, experiencing Treatment A prior to Treatment B results in an increase of 6 (i.e., +6), whereas all other sequences result in an increase of 2 (i.e., +2).

	Sequence of Treatments		
	A (1 M&M)	B (3 M&Ms)	C (5 M&Ms)
Effect on Performance →	(0	+6	+2)
	A	C	B
	(0	+2	+2)
	B	A	C
	(0	+2	+2)
	B	C	A
	(0	+2	+2)
	C	A	B
	(0	+2	+6)
	C	B	A
	(0	+2	+2)

drastic decrease in performance produced by the C to A (five M&Ms to one M&M) transition is not canceled out by a comparable increase resulting from the A-to-C transition.

The potential for differential carryover exists whenever counterbalancing is used. The experimenter must be sensitive to this possibility. A thorough review of the effects of the IV being used in your research can help sensitize you to the possibility of differential carryover. If this threat exists, it is advisable to employ a research procedure that does not involve presenting more than one treatment to each participant.

REVIEW SUMMARY

1. **Randomization** controls for extraneous variables by distributing them equally to all groups.
2. When experimenters use **elimination** as a control technique, they seek to remove completely the extraneous variable from the experiment.
3. **Constancy** controls an extraneous variable by creating a constant or uniform condition with regard to that variable.
4. **Balancing** achieves control of extraneous variables by ensuring that all groups receive the extraneous variables to the same extent.
5. **Counterbalancing** controls for sequence or order effects when participants receive more than one treatment. **Within-subject counterbalancing** involves the administration of more than one treatment sequence to each participant, whereas **within-group counterbalancing** involves the administration of a different treatment sequence to each participant.
6. The total number of treatment sequences can be determined by *n*! When all sequences are administered, **complete counterbalancing** is being used. **Incomplete counterbalancing** involves the administration of fewer than the total number of possible sequences.
7. The random selection of treatment sequences, systematic rotation, or the Latin square approaches can be used when incomplete counterbalancing is implemented.
8. Counterbalancing can control for **sequence or order and carryover effects**. It cannot control for **differential carryover**, in which the response to one treatment depends on which treatment was experienced previously.

Check Your Progress

1. Matching

1. randomization	A. complete removal of the extraneous variable
2. elimination	B. extraneous variable is reduced to a single value
3. constancy	C. most widely used control procedure
4. balancing	D. used to control for order effects
5. counterbalancing	E. extraneous variable is distributed equally to all groups

2. The most widely used control technique is
   a. randomization
   b. elimination
   c. constancy
   d. balancing
   e. counterbalancing

3. Balancing is a logical extension of
   a. constancy
   b. randomization
   c. counterbalancing
   d. elimination

4. Distinguish between within-subject and within-group counterbalancing.

5. What can *n*! be used for? Calculate the value of *4*!

6. ________ occurs when the response to one treatment depends on which treatment preceded it.
   a. Carryover
   b. Differential carryover
   c. A sequence or order effect
   d. Fault randomization

7. What is incomplete counterbalancing?

Key Terms

Variable, 100
Operational definition, 101
Physiological IV, 102
Experience IV, 102
Stimulus or environmental IV, 102
Participant characteristics, 103
Confounding, 105
Valid, 108
Reliable, 108
Nuisance variable, 109
Randomization, 113
Elimination, 114
Constancy, 114
Balancing, 115
Counterbalancing, 117
Within-subject counterbalancing, 117
Within-group counterbalancing, 117
Complete counterbalancing, 119
Incomplete counterbalancing, 119
Sequence or order effects, 121
Carryover effect, 122
Differential carryover, 123

Looking Ahead

In this chapter we have explored the nature of variables in general and have seen the importance of selecting appropriate IVs and DVs. The potentially damaging effects of nuisance variables and confounders have led to the development of procedures for their control. We continue our examination of the basics of experimentation in the next chapter. There we will discuss the selection of appropriate types and numbers of participants and the actual collection of research data.