Chapter 9. Using Experimental Control to Reduce Extraneous Variability

Introduction to Experimental Control

In the previous chapter, we saw how extraneous variables can contribute to systematic error and to random error. Both sources of error reduce the internal validity (quality) of research and make interpretation of results difficult. As we discussed issues such as demand characteristics, experimenter bias, and testing effects, you likely had some ideas regarding how we might design a study to reduce or eliminate these and other sources of error. Indeed, the researcher has many research design tools that can be used to address these concerns. The good researcher is aware of both the sources of extraneous variability and the research design techniques that can be used to reduce or eliminate extraneous variables.

Characteristics of a True Experiment

In this book, we explore a variety of research methods. In fact, for most areas of research, we have a choice of several research methods. These methods incorporate techniques designed to reduce or eliminate sources of extraneous variability and permit more powerful conclusions. Through the next several chapters, we focus on the most powerful methodology available—the true experiment.

Advantages

The hallmark of the true experiment is control. The experimenter is in control of many facets of the research design. The experimenter controls the way in which a sample of participants is obtained from the population, participants are assigned to different treatment conditions, the environment is organized during testing, instructions are presented to participants, observations are made, and data are collected. As we will see, the purpose of this control is to reduce the influence of extraneous variables so that changes in the dependent variable can be attributed to the independent variable.

In this chapter, we will describe random assignment, the use of control groups, and careful experimental techniques as means of reducing extraneous variability and increasing internal validity. In brief, we will look at how research should be done. Keep this principle in mind: The time to avoid random error (the largest component is individual differences) and confounding (systematic error) is during the design phase. Possible sources of confounding should be anticipated and eliminated before gathering data. After the data have been gathered, it is too late to eliminate any confounding that may exist.

Limitations

Generally speaking, whenever a true experiment can be used to answer a research question, it should be used. However, experiments have their limitations. There are situations in which a true experiment cannot be used because of the nature of the research question or because of ethical issues. As discussed in the

previous chapter, a potential limitation of many experiments is the generalizability of the findings from the controlled environment of the laboratory to the real world. That is, can the pattern of results in the experiment be extended to behavior in the real world? To achieve the highest degree of experimental control, most experiments take place in a laboratory setting. To what extent will this artificial environment and the particulars of the setting contribute to the patterns of behavior that are observed? For example, is learning a list of words in a laboratory the same as learning in the classroom?

Some research questions do not lend themselves to the true experiment. The question "Are men more aggressive than women?" cannot be answered using a true experiment because random assignment to the two groups cannot be used. The same issue applies to a host of physical and personality characteristics (for example, anxious versus nonanxious, heavy versus thin, optimist versus pessimist). Consider another example. We might ask whether a change in the level of violence on U.S. television in the past 50 years has corresponded with a change in the level of juvenile delinquency in the United States. This is a valid question, but obviously the events have already occurred (as in our men/women example). We cannot randomly assign participants to treatment conditions and thus cannot achieve the level of control required for a true experiment.

Another example will highlight situations in which ethical concerns are paramount. Let's say that we are interested in whether children who are abused are more likely to grow up to become parents who abuse their own children. This is a very interesting and important question, but not one that can be answered with a true experiment. Can you imagine the following experimental procedures? We go out and sample a group of children to be in our study. We then randomly assign some of them to be physically abused for the next five years, some to be sexually abused for the next five years, some to be neglected for the next five years, and some to be in families with wonderful parents. We then wait for these children to grow up and have their own children, and we observe their parenting behaviors. Might there be an ethical issue here?

Thus, the true experiment cannot be used in every research situation, because of either methodological or ethical issues. We dedicate a chapter toward the end of this book to alternative research designs. But for the moment, we focus our attention on the characteristics of a true experiment and the procedures for conducting a true experiment.

The Notion of Experimental Control

The previous chapter introduced us to the variety of factors that can be sources of extraneous variability. As we noted, the challenge of good research is to develop a research design that will eliminate or reduce sources of systematic error and random error so that systematic variance resulting from manipulation of the independent variable can be revealed. The primary goal is to create a research design that is internally valid. Achieving this goal may seem rather difficult, and it is true that there is no perfect experiment.

However, behavioral scientists have a set of methodological tools at their disposal that can be very effective in controlling sources of extraneous variability. The key is to understand what is in the toolbox and when to use it. Figure 9.1 summarizes the control tools available. Let's open this toolbox and take a peek inside.

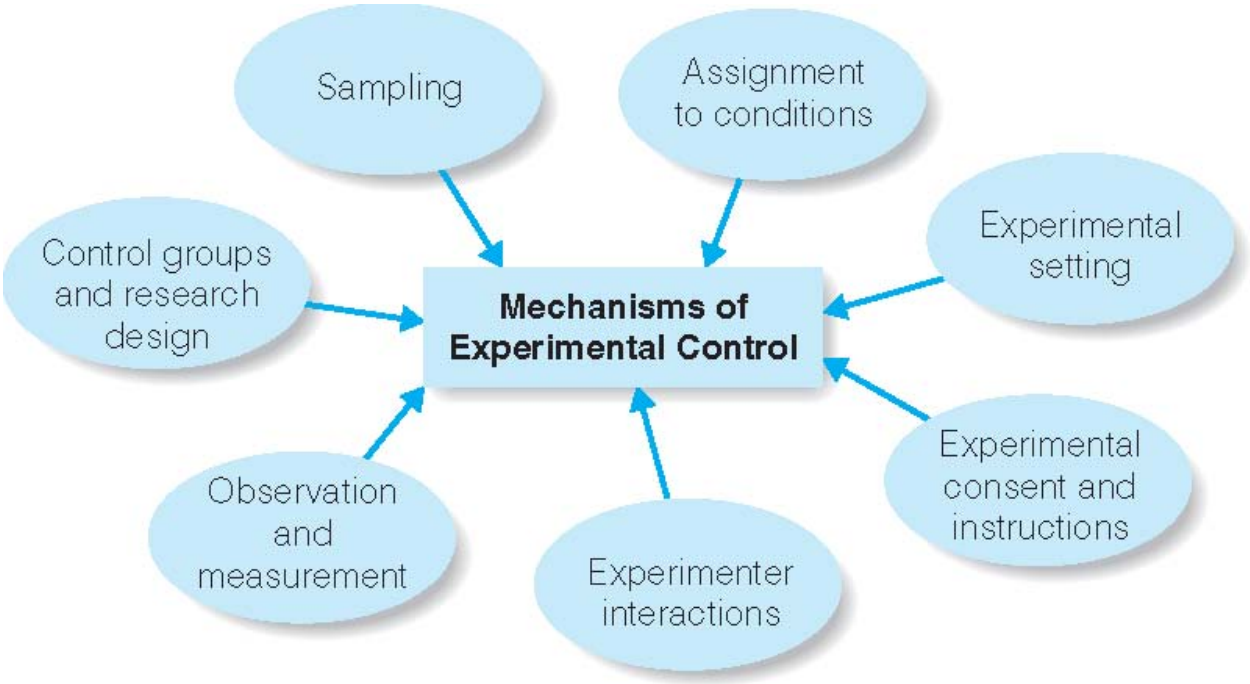


Figure 9.1 Summary of the research design tools that are available to achieve experimental control.

Control Through Sampling

Methods of sampling, discussed in Chapter 7, can effectively reduce extraneous variability due to selection and regression to the mean. Remember that we want to select a sample of participants that is representative of the population of interest. Also, we need to realize that samples selected for their extreme scores on some variable will likely show a drift in their scores away from the extremes—that is, toward the mean. As discussed previously, random sampling is often the best approach to obtain a representative sample. Random sampling not only controls several extraneous variables, it also allows us to generalize to a given population (increases external validity). However, as we discussed in Chapter 7, it is often difficult to obtain truly random samples, and the researcher often resorts to the use of convenience sampling (such as using introductory psychology students). Thus, the researcher should consider how the sampling procedure might affect the study's outcome when truly random sampling is not possible.

Control Through Assignment to Conditions

Although we have not yet discussed specific research designs, it is clear by now that experiments typically involve a comparison of scores obtained under different conditions. Methods for assigning participants to conditions can control a variety of extraneous variables. These methods fall into two categories. The first category involves the creation of groups by random assignment. This technique creates what is termed **independent samples,** and it is the best way that we know to create equality of groups on all known and unknown factors. The second category involves the creation of correlated samples by pairing scores. We provide a brief explanation of each category here; specifics are provided in subsequent chapters.

Independent Samples Design

Random assignment relates directly to internal validity and is concerned with the way in which we assign participants to experimental conditions. It is an essential characteristic of experimentation. The purpose of random assignment is to avoid bias in the composition of the different groups. We want to create groups that are essentially equal so that any differences we subsequently find can be attributed with some confidence to the effects of the treatments themselves, assuming that everything else is held constant. We want to be reasonably sure that the independent variable, and not the method of assigning participants to groups, gave rise to the obtained differences. Random assignment is the best way of doing this. Moreover, random assignment of participants to experimental conditions is a basic assumption of many statistical techniques that we use to make inferences from samples to populations. Satisfaction of this assumption is essential for using these statistical procedures.

With **random assignment,** groups (independent samples) are created such that each participant has an equal chance of being selected for a particular experimental condition. Although random assignment does not guarantee the formation of equal groups, it is the best way that we know to create equality. The value of this technique can be demonstrated by considering the opposite technique—asking for volunteers. If we ask our sample of children in our TV violence study, "Who would like to watch *Beast Wars*, and who would like to watch *Mister Rogers*?" you can imagine the extraneous variables likely to surface, and you would likely observe these variables in the children's response to the question. In response to the request for a *Beast Wars* group, you would likely observe a group of predominantly boys emphatically thrust their hands in the air and yell "Oh, yeah!" Conversely, the volunteers for *Mister Rogers* might meekly raise their hands to half extension and quietly reply "I'd like to." Although this may be a bit of an exaggeration, the point is clear. By not using random assignment, you run the risk of creating groups that are systematically different on a host of variables before the independent variable is even manipulated.

Correlated Samples Design

Control can also be achieved by pairing scores to create **correlated samples.** The paired scores in the groups may represent natural pairs, matched pairs, or repeated measures.

Natural pairs. In a **natural pairs** design, the scores in the groups are paired for some natural reason. Studies using twins provide common examples in the behavioral sciences. One typical source of extraneous variability is that the participants in the comparison groups have different genetic backgrounds, and this factor contributes to the random error in the scores. One solution is to equate the genetic backgrounds by using identical twins. For each pair of twins, one twin is randomly assigned to one condition, and the other twin is assigned to the other condition.

Matched Pairs. In a **matched pairs** design, the scores in the groups are paired because the experimenter decides to match them on some variable. In some experiments, the experimenter may decide that there is some extraneous variable so critical to the research that the researcher does not want to rely on random assignment to equate the groups on that variable. For example, if our sample of children for our TV violence study contains children of various ages, we may decide that age is an extraneous variable that must be equated across the groups. Thus, we begin by grouping the children in our sample according to age. We then randomly assign them to the two groups according to age. That is, if we have four 5-year-olds in our sample, two will be randomly assigned to one group and two to the other group. Therefore, when we complete data collection and compare scores in the two groups, we are assured that age is not contributing to the variability in scores between groups.

Repeated Measures. In a **repeated measures** design, the scores in the two groups are paired because they come from the same participants. In other words, each participant is tested in each experimental condition. This technique can be an excellent method to reduce random error between groups that is due to individual differences. For our TV violence study, we could test all the children under both conditions. That is, one week the children could be presented with a TV program with violence and their behavior observed, and another week the same children could be presented with a TV program without violence and their behavior observed. However, the method does raise the possibility of carryover effects—an issue that is discussed more fully in Chapter 12.

Control Through Experimental Setting

Experimental control is enhanced by selection of a setting that you can control. That is, you can control the size, temperature, and location of the setting. You control when participants enter the setting, where they stand or sit, what they see, hear, smell, taste, and touch. In other words, you control the environmental stimuli that they experience.

Let’s return again to our TV violence example and assume that we are comparing level of aggressive behavior in a group of children who watch a TV program with violence to a group of children

who watch a TV program without violence. We can control the experiment setting in a number of ways. We can designate a particular room in the psychology building for testing. We can decide on the size of the room, the color of the walls, the size of the TV, the volume setting on the TV, the type and amount of furniture in the room, the type and number of toys that may be available for the observation period, the number of children in the room during testing, the number of experimenters in the room during testing, the duration of testing, and the time of day of testing. Notice that each of these factors is a potential source of extraneous variability. We want to carefully consider each factor and be sure that both groups of children have the same experimental setting (except for the type of TV show).

In making these decisions, it is important to consider both internal and external validity. In addition to designing an experimental setting that maximizes control of extraneous variables, we also want to design a setting that seems real, so that we will be more confident that our findings will generalize to the real world. Internal validity might be enhanced by testing the children in a room that is completely empty except for a TV, but we would likely sacrifice some degree of external validity.

Meticulous attention to all aspects of our experiment must be pursued right down to the administration of the treatments to the participants. For example, it might be convenient for you to give the treatment to all participants in the experimental group at the same time (all at once) in one session, and to participants in the control group (all at once) in a different session. However, there are risks associated with this procedure. Under these conditions, the individual scores may not be independent. It is possible that any extraneous or unwanted event that occurs during that session could have a marked effect on the performance of one of the groups and not the other.

In essence, we risk confounding the “time and setting” of administering the treatment with the treatment itself. Here are some examples: one very unhappy and uncooperative or unruly participant in one group and not in the other; a hot, noisy room versus a cold, damp room; a knowledgeable, helpful experimenter versus one less knowledgeable and helpful. Some participants may complete the task quickly in one condition and create panic among others in the group who perform more slowly; or the groups may interact differently and ask different questions regarding the task and the instructions, thus resulting in groups very different from what the experimenter envisioned. In short, many things may happen that affect all participants in a particular condition. The scores of individual participants would no longer be independent. Thus, careful thought must be given when choosing between a procedure that administers treatments individually and one using intact groups. Obviously, if intact groups are to be tested, great care must be taken to see that each group is treated as similarly as possible. An alternative to intact groups would be to test several smaller groups under a given condition or to test one individual at a time.

If we decide to test participants one at a time rather than in intact groups, then other considerations arise. To start with, good experiments have, at a minimum, two comparison groups (usually an

experimental group and control group) to which participants are randomly assigned and then tested individually (one at a time). There is a proper procedure that should be followed when testing participants one at a time. On any given day, an equal number of participants from each condition (experimental and control groups) should be tested, so that each treatment is represented daily. In an experiment with two conditions, at least two participants (or multiples of two) should be tested on any given day, one from each condition; in an experiment with three conditions, at least three participants (or multiples of three) should be tested on any given day, one from each condition. It is not a proper procedure to administer the treatment to all participants individually in one treatment condition first and then to all participants in another treatment condition. If the latter procedure were used, systematic changes that occur in the separate experimental sessions, or between the time when the first treatment was presented (Time 1) and when the second treatment was presented (Time 2), would be mixed with the effects of the independent variable—that is, confounding. As experimenters become more experienced in dealing with participants, apparatus, instructions, and data recording, they change in some ways. If, in an experiment with two treatment groups, one group was tested first and the other tested second, the experimenter may be naïve for the first group but knowledgeable and sophisticated for the second group. Further, the experimenter may be in a different psychological state (bored) or physical state (ill) for one group. These changes between Time 1 and Time 2 could result in the experimenter's treating the two groups differently.

Other events capable of exerting a systematic effect could occur between Time 1, when the first group is tested, and Time 2, when the second group is tested. Participants in one group may be getting the treatment during midterm or final examinations; the other group, at a different time. Measuring instruments, clocks, and other equipment may become less reliable; observers may change their scoring criteria. The important point is that the procedure of testing one group or treatment condition first, either all at one time or individually, is a faulty one and should be avoided.

The following principle should be kept in mind when planning your experiment: If any extraneous variable with the potential for exerting a systematic effect cannot be eliminated, then it must be held constant for each treatment group—that is, its effects must be distributed to each treatment group as equally as possible. Balancing participants so that each condition is equally represented each day automatically takes into consideration possible variables such as time of year, seasons, and time of school term (such as midterms or finals). When more than one participant is tested each day, balancing for the time of day that they are tested is also necessary. We should note, also, that if more than one experimenter is involved in collecting data, each should test an equal number of participants under each condition.

Control Through Experiment Consent and Instructions

With human participants, careful consideration of the communications to participants via the consent form and instructions can control for such extraneous variables as demand characteristics, evaluation

apprehension, diffusion of treatment, task, and instructions. The language used should be relaxed and professional. Our goal is to observe natural behavior. We want the participants to feel relaxed and not experience apprehension about being evaluated or observed.

Although the consent form should provide sufficient information for informed consent, unnecessary details and specific hypotheses should be avoided. Such details can quickly lead to demand characteristics on the part of participants. If the participants are to be assigned to a particular group, it is often wise to use a **single-blind study** in which participants do not know to which group they have been assigned. Again, knowing which group they are in can create demand characteristics. This can be a particular problem in drug studies, where belief in the efficacy or effects of a drug can cause changes in behavior regardless of the actual effect of the drug. These placebo effects and the use of placebo control groups will be discussed a little later in this chapter.

It is critical that the same instructions be provided to every participant in the study (unless the instructions themselves are the independent variable). This can be accomplished in several ways. Instructions can be read from a script or presented as an audio recording. Participants may read the instructions on a sheet of paper or on a computer screen. If diffusion of treatment is a concern, then it is useful at the end of participation to instruct participants not to discuss the experiment with others who might be future participants.

As we discussed in Chapter 2, the use of deception is sometimes necessary to control demand characteristics so that natural behavior is observed. A good example is the classic study on conformity by Solomon Asch (1956). Asch was interested in whether a participant, judging the length of lines, would provide an incorrect response simply because several others had done so (results showed that they often did). If participants had been informed that this was a study of conformity rather than a study of visual perception, it is unlikely that natural behavior in a situation with social pressure would have been observed.

Let’s consider the instructions that might be given to the children in our TV violence study. Of course, it would be important to give both groups of children exactly the same instructions. What would we tell them when they arrived for the experiment? We would probably want to use the single-blind technique and provide general instructions without specifics. For example, we might tell them that they are going to watch a TV program and that when the program is finished, they will be able to play in the room. Notice that the children will not know which group they are in and will not know that aggressive behaviors will be recorded during the play period. We suspect that you would have thought of this. However, have you considered what the experimenter would do if one or more children stopped attending to the TV program, if one or more children wanted to see a parent before the testing was complete, or if one or more children hurt another child? Would you intervene? If so, what would you say? These types of

situations are the ones often overlooked during the planning stages of the experiment. The key is to develop a protocol for every possible situation that you can imagine.

Control Through Experimenter Interactions

We have mentioned several ways that the experimenter can be a source of extraneous variability, including experimenter bias in observations, experimenter effects, enhancement of demand characteristics, and enhancement of evaluation apprehension. One additional dimension that deserves attention is the professional demeanor of the experimenter during interactions with the participants. The importance of this factor should not be underestimated. The quality of data obtained from human participants is directly related to the seriousness with which they assume their role as a research participant. If the experimenter is dressed unprofessionally, appears unprepared, or jokes around with participants or other experimenters, then the participants are less likely to take their participation seriously. They are less likely to follow instructions, attend to stimulus presentations, and do their best. All efforts to design a high-quality experiment can be wasted if the experimenter acts unprofessionally.

Control Through Observation and Measurement

Methods of observation and measurement have been discussed in previous chapters. Particular methods can effectively reduce extraneous variability due to demand characteristics, evaluation apprehension, experimenter characteristics, experimenter bias, and instrumentation. Recall that observations can take place with or without the participant's awareness. If observation without awareness is used, demand characteristics and evaluation apprehension are greatly reduced. For example, in some psychological experiments, the critical observations have taken place as participants waited in a waiting room. Similarly, if children are observed without their knowing that they are being observed, more natural behavior is expected.

In the previous examples, participants were unaware that they were being observed. In situations where the participants know that they are being observed but the observer is not physically present, one is more likely to observe natural behavior and to eliminate extraneous variability due to experimenter characteristics. If the experimenter is not present, characteristics of the experimenter (such as gender, age, or attractiveness) cannot affect participant behavior. This can be accomplished, for example, by using one-way mirrors, video recording, or computer-controlled protocols.

We noted earlier in this chapter that participants can be affected by knowing which group they are in and that a single-blind technique can be used to control this problem. Likewise, experimenter observations can be affected by knowing which group a participant is in. In the previous chapter, we referred to this potential confound as experimenter bias. It can be controlled by making the experimenter

"blind" to the condition to which the participant was assigned. An experiment in which neither the participant nor the experimenter knows which group the participant is in (at least when the observations are made) is referred to as a **double-blind study.**

Another issue already discussed is the use of specific operational definitions and multiple observers. Both reduce the opportunity for experimenter bias and provide a mechanism for high interobserver agreement. Finally, instrument decay should be avoided by verifying that the actual recording of data has not changed over time or varied with the experimental conditions. This may involve periodic checks on observer performance (observer drift) or periodic calibration of equipment.

Let's consider the information discussed so far in the chapter by applying it to a news report regarding the effect of lead exposure in children (see "Thinking Critically About Everyday Information").

Thinking Critically About Everyday Information: Effects of Exposure to Lead

ABC News reported new research on the effects of children's exposure to lead on development. A portion of the report on their Web site follows:

> Poisoned Minds: Lead Levels Linked to Lower IQ in Children
> . . . The harmful effects of lead poisoning on children are well-documented, but new research suggests that the danger is more widespread than ever imagined, and that exposure to levels currently deemed safe can lower children's IQ scores. . . . A study recently published in the *New England Journal of Medicine* looked at the blood lead levels of 172 children in Rochester, N.Y., ranging in age from six months to 5 years. Researchers tested the children's IQ at ages 3 and 5, and found that those whose blood levels of lead increase from one microgram per deciliter to 10 (the limit under CDC's safety guidelines) experienced an IQ drop of 7.4 points. Children whose blood levels rose from 10 to 30 micrograms per deciliter lost an additional two to three IQ points. But the key point in the research is that even at levels below the limit deemed safe by the CDC, children were losing IQ points. A separate study in the journal from the Environmental Protection Agency found that low levels of lead delay puberty for several months in young girls, especially African-Americans and Latinas. The concern is that the lead is interfering with hormonal processes during development. In addition, a University of Pittsburgh study found that juvenile offenders had a much higher concentration of lead in their bones compared to their counterparts who were not in trouble with the law.

These are important and interesting findings. Now think about how these studies were done. To this point in the chapter, we have discusses characteristics of true experiments and several control techniques that can be used.

- Do the above studies qualify as true experiments? Why or why not?
- Was the setting controlled by the experimenter?
- Was there random assignment to different lead exposure levels?
- Based on your answers to these questions, what can you conclude from the above studies? What can you not conclude?

SOURCE: Retrieved June 10, 2003, online at
http://abcnews.go.com/sections/GMA/AmericanFamily/GMA030602Lead_level_research.html

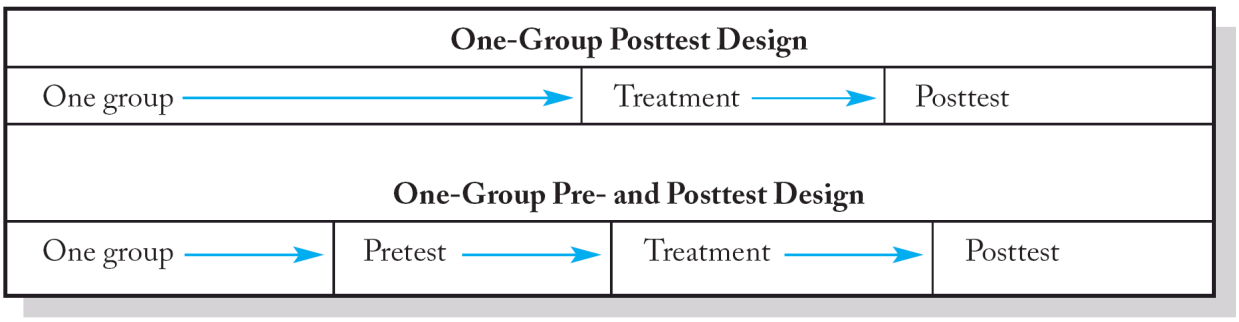
Control Through Use of Control Groups / Research Design

True experiments involve making comparisons between participants' scores in different experimental conditions. Thus, the design of these comparison groups and the way in which participants are assigned to these conditions are critical in determining the integrity of the research design. If comparison conditions are not included in the research design or the assignment of participants to conditions does not follow established methodological guidelines, then the extraneous variables discussed in the previous chapter make interpretation of the data impossible.

Primitive Research Designs

To highlight this last point, we will begin by discussing two primitive research designs that do not meet the standards of a true experiment. These two primitive designs give rise to the concept of a control or comparison group. Without proper control conditions, results are generally uninterpretable, and the research is often useless. With a primitive design, it is virtually impossible to determine whether the relationship is between the dependent and independent variables or between the dependent variable and some unwanted variable. With this type of design, it would not matter how carefully the observations were made; the data would remain uninterpretable.

Two primitive designs that are still occasionally used are the one-group posttest design and the one-group pre- and posttest design. These two designs are depicted below.



With the one-group posttest design, a single group of participants is selected, a treatment is given, and the behavioral effects of the treatment are measured (posttest). For example, we could sample a group of children, expose them to TV shows with violence for one week, and then measure their level of aggressive behavior the following week. Let's assume that our group of children exhibited a mean of four aggressive behaviors in the 60-minute observation periods. What does this tell us? Did the violence in the TV programs cause these aggressive behaviors? If your answer is "there is no way to know," then you are thinking like a researcher. Of course there is no way to know. The most obvious shortcoming is that there is no comparison condition. There is nothing to compare their aggressiveness to. How aggressive are these children normally? The children may have exhibited the same level of aggression (or even a

greater level) without the TV programs. There are various other problems with this design that we will not dwell on. Suffice it to say that it is primitive, lacking many features of better designs.

The one-group pre- and posttest design is an improvement over the first design because it permits us to say at least some things. In this case, we assess the children's level of aggressive behavior before viewing the TV programs and after viewing the programs. In contrast to the one-group posttest design, we now have a standard (pretest) against which to compare any changes in aggression that occur on the posttest. Let's imagine that we find a marked increase in aggression between the pre- and posttest scores. Can we attribute the change to our independent variable—that is, our presentation of the TV programs with violence? Or are alternative accounts also possible? Again, you are thinking like a researcher if you concluded that the second design does not rule out alternative accounts.

Other factors could have occurred during the period when the children watched the TV programs with violence that are fully or partially responsible for the change in aggressive behavior. However, their effects are confounded with the effects of the TV programs. Events at home, at school, or in peer groups could have affected the children's behavior from one week to the next. Also, the natural maturational process could be responsible.

The one-group pre- and posttest design does not provide a way to assess such historical or maturational factors. Indeed, the effects of testing itself may be a factor and should be considered as a possible basis for any change in behavior that is observed. In our example with the children, there is no test per se, but there is an initial observation period that may sensitize the children to being measured and sensitize them to the types of issues involved. These issues are particularly relevant when it comes to an actual test or questionnaire. Students taking the same or similar tests a second time often score differently. The pretest may sensitize them to the kinds of issues involved, and they may react in ways that are unpredictable. Questions on the pretest may make the participants more aware that issues exist. After taking the pretest, they may decide that certain answers are more socially desirable on the posttest. Further, participants may become more cooperative and trustful regarding the questionnaire only after experiencing the pretest. Whether or not these things actually happen is not at issue. The point is that the possibility exists. Consequently, we are unable to untangle the effects of the independent variable from the "spaghetti" of possible alternative explanations.

In some instances, our examples of alternative accounts may be weak and debatable. We do not want you to focus on this point. The important issue is that we cannot assess separately the possible effects of the independent variable and the possible effects of these other factors. Other problems associated with the one-group pre- and posttest design may include such extraneous variables as regression toward the mean, demand characteristics, participant expectancies, and experimenter bias.

A point should be made that some of the criticisms of this design are not valid when the experiment is very short term and takes place under laboratory conditions in which the participant is relatively

isolated. In short-term laboratory studies, few events are likely to occur between pre- and posttests, nor are major changes in the participant likely to occur. However, test sensitization and the other factors noted above continue to play an important role.

Importance of Control Groups

A simple addition to the two designs discussed so far would improve the research method considerably. This simple addition would rule out a number of the alternative interpretations that we noted were possible with these designs. Adding a control or comparison group that did not watch TV programs with violence would provide valuable additional information. The comparison group would have to be similar to the treatment group (experimental group) and treated in an identical manner on the pretest and posttest. The only difference is that they would not watch programs with violence. Under these circumstances, we could better isolate the effects of our independent variable. If the groups are assigned randomly and treated properly, we could eliminate most of the alternative accounts noted for the primitive designs.

The use of control groups is important in both laboratory and applied settings. This is particularly the case when new techniques of therapy are being evaluated. Often, when a new medical or psychological therapy is introduced to a group of patients, they show a remarkable recovery. It is tempting to conclude that the improvement is due to the treatment. However, this conclusion cannot be supported without a control group. Recovery from the disorder may not be due to the treatment—medical therapy or psychotherapy—but may arise from other factors. Some individuals may have recovered spontaneously. We are reminded of the frequent observation about the common cold: With the finest of medical treatment, it will be “cured” in a week; otherwise, the patient will require seven days to recover. It is also possible that the simple act of giving attention and showing concern for the well-being of the individual may be the important factor (placebo effect). Participants may expect to get better because experts are attending to them, technology is being used, and gadgets are on display—more placebo effects. Alternatively, experimenter expectancies may operate to the extent that the experimenter sees improvement when there is none. Control groups combined with a single-blind or double-blind procedure would be essential in this situation.

Sometimes data derived from studies without proper control groups are very compelling, even though alternative accounts are plausible. These data do not have to be dismissed. They should serve as a basis for properly designing a study. Let’s consider an example dealing with blood dialysis for schizophrenics.

Researchers continue to learn more about the causes of schizophrenia. One view is that the disorder is linked to a chemical imbalance in the brain. About 20 years ago, a controversial study was reported that relates to this biochemical view. A University of Florida professor of medicine, after reviewing the evidence on schizophrenia, believed that a strong case could be made that it was an inherited disorder. He then assumed that it had an organic basis and that it might possibly be related to the individual’s blood

supply. If this was so, then it might be possible that the material circulating in the blood could be removed by dialysis. Dialysis is a process used to remove waste material from the blood of patients with kidney disorders. The researcher decided to try this "blood cleansing" process on schizophrenics. He selected 16 patients diagnosed as schizophrenic and treated them with dialysis. The dialysis treatment he used was the traditional, well-established procedure (the particulars are unnecessary), and the patients were fully informed of the nature of the treatment. His results were remarkable. Hallucinations and depression disappeared in 14 of the 16 patients. These same patients also showed a considerably improved ability to adapt to normal social situations—often a difficult ordeal for schizophrenics. Could it be that our investigator had discovered a cure for schizophrenia?

Results may have been favorable because patients wanted to believe in the treatment, and in a therapeutic/caring environment their expectations may have been very high. Also, the therapist (researcher) evaluating the subsequent behavior of the schizophrenics was aware that they had undergone dialysis treatment. In this case, the experimenter's expectancies, if any, could be biasing the outcome. The question could also be raised regarding how the patients were treated by others during this period of receiving dialysis. If they were treated in a different way, then this is another form of treatment different from dialysis. The results were very dramatic and certainly suggested that follow-up research on the new technique be undertaken with a much better design.

To rule out expectancy effects, experimenter bias, other forms of inadvertent treatment, and other possibly important factors, a control-group design using random assignment is needed. In contrast to the one-group pre- and posttest design, this design rules out a large number of alternative interpretations.

All schizophrenic patients would first be evaluated (pretested) by a group of independent therapists not involved with the research. The patients would be randomly assigned to the placebo control condition (control for expectancy plus other factors) or to the experimental condition. The patients would be "blind" in the sense that they would not know what condition they were in. Those in the control group would go through a "dummy dialysis" procedure, in which blood would be drawn but not treated. In other words, all patients would be treated exactly alike except for the actual dialysis given the experimental group. Only an attending physician would know who received the treatment. This procedure would assure that the therapists evaluating the subsequent behavior of schizophrenics were also "blind," thus avoiding therapist (experimenter) bias. Because both patients and experimenters would be blind, the study is described as a double-blind one. To avoid changes in our instrumentation, the same therapist would evaluate the patients pre- and posttreatment. At the end of the first phase, all patients would be evaluated and the second, crossover phase begun. In the second phase, the experimental and control participants would have the conditions reversed. Those who were in the experimental treatment would have it withdrawn, and those in the control condition would have the treatment introduced—that is, the groups would cross over to the other condition. No information would be given at this time, and participants

would be unaware that a change had taken place. It goes without saying that before the experiment began, all participants would be informed of the two conditions of the experiment and asked to consent to participate.

Soon after the initial report of success with dialysis, Schulz, van Kammen, Balow, Flye, and Bunney (1981) studied eight chronic schizophrenic patients using a double-blind procedure. None of the patients improved with dialysis; in fact, four patients got worse. To this day, dialysis is not considered a treatment for schizophrenia.

Designs With Control Groups

The importance of control groups cannot be exaggerated. When properly used, they allow us to isolate the effects of the independent variable. In doing so, we can distinguish the effects of the independent variable from the effects of other variables that might produce systematic error. The proper control groups allow us to conclude that the observed relationship is between our independent and dependent variables.

As we noted, the primitive research designs lack control groups, and thus the data they provide, though suggestive, are not interpretable. We will now describe two experimental designs that make use of random assignment and control groups. These are powerful designs for isolating the effects of the independent variable. The randomized posttest control-group design and the randomized pre- and posttest control-group design are true experimental designs that use randomization to assign participants to the experimental treatment conditions and to the no-treatment control condition. The only difference between the two designs is the presence or absence of a pretest.

Randomized Posttest Control-Group Design

Group 1 →	Treatment →	Posttest
Group 2 →	No Treatment →	Posttest

Randomized Pre- and Posttest Control-Group Design

Group 1 →	Pretest →	Treatment →	Posttest
Group 2 →	Pretest →	No Treatment →	Posttest

The simplest instance of a control-group procedure can be seen in these two designs. The treatment and no-treatment groups are treated exactly alike except that one receives the independent variable, and the other does not. A comparison (baseline) condition is provided by the no-treatment control group to judge the effects of the treatment condition in the experimental group. We should add that the term *no treatment* should not always be taken literally. It simply means that the experimental treatment was not

given to this group. At times, some form of treatment different from the experimental treatment is given. On other occasions, several groups are used, each receiving lesser amounts of the independent variable. Designs involving systematic variations in the amount of a given independent variable, known as parametric designs, will be discussed in Chapter 13.

Experimental designs with control groups are powerful. The addition of a control group combined with random assignment greatly increases our understanding of what is occurring in our experiment. As noted, if a relationship is found, these designs allow us to rule out many competing interpretations. Therefore, with these designs, we can separate the effects of our independent variable from such factors as individual histories, maturation, participant selection, testing effects, participant expectancies, regression effects, and various other possibilities.

Because random assignment is used in these designs, it is likely that the two groups are about equal and that any differences observed on the posttest measure between the two groups are due to the independent variable and not to participant bias. In this regard, the use of a pretest has both positive and negative facets. Researchers sometimes use a pretest measure to assure themselves that the random assignment procedure did, in fact, result in equivalent groups before treatment. This assurance is not usually necessary; most researchers believe that random assignment in itself is sufficient.

Pretests can be useful. A pretest measure makes it easier to assess whether any loss of participants (attrition) during the course of the experiment results in a bias for one group (inequality between groups in terms of participant attributes). Determining whether attrition has biased one of the groups can be done easily with pretest scores by assessing the scores of those remaining in the experiment in the two groups or by comparing those lost from the experiment for the two groups. Pretests also permit us to use a more powerful statistical test such as analysis of covariance, given that the assumptions for its use are met. A covariance analysis is a finely tuned statistical instrument that permits the detection of small treatment effects. Pretests are also useful, sometimes necessary, if you want to assess how effective a treatment is for a specific individual—for example, determining the level of anxiety before therapy and again after therapy.

At times, a pretest measure may not be available, or you may be concerned with the potential problems it might create. Pretest measures require more time, effort, and expense. Moreover, they are often inconvenient for both participant and experimenter. Recall that they may also sensitize the participants, causing some to form hypotheses concerning the experiment (demand characteristics). If the purpose of a pretest is to determine whether random assignment has resulted in equal groups, it can be dispensed with; the design with only a posttest would be as effective. However, if participants cannot be assigned randomly, as in quasi-experimental designs (discussed in Chapter 11), pretest measures are essential. Without these measures, researchers would have little idea about the equivalence or non-equivalence of their groups.

Obviously, these two experimental designs have advantages and disadvantages, depending upon the needs of the experimenter. These relate to the pretest and whether or not it serves some essential function other than assessing the equivalence of groups. We now turn to more detailed discussions of two specific types of control groups: yoked control and placebo control.

Yoked Control Procedure. The **yoked control** procedure allows the researcher to isolate the important effects of the independent variable while holding other possible factors constant. Perhaps you recall reading about the "executive" (experimental) monkeys and the "employee" (control) monkeys described in a study by Brady (1958). Using a yoked control design, Brady placed experimental and control monkeys in identical restraining apparatus where mild shock was periodically delivered. The only difference was that the experimental executive monkeys could avoid shock by responding properly. They learned the task quickly but not perfectly, and on some occasions they received shock. Whenever the executive monkey received shock, the employee monkey, yoked to the executive, also received shock. Both experimental and yoked participants thus received the same number of shocks, the same duration of shocks, and the same temporal distribution of shocks within each experimental session. The only difference between the participants was that for one the response–outcome relationship was dependent (contingent—participant has control) and for the other the relationship was independent (noncontingent—participant has no control).

The executive monkeys, who could control whether or not shock was avoided, all developed serious ulcers, but the employee monkeys did not. Brady concluded that ulceration was a function of the executive role forced upon the monkeys. This conclusion is intuitively satisfying to many people, and belief in the debilitating effects of the executive role is widespread. Yet subsequent studies, such as those of Weiss (1968, 1971a, 1971b, 1971c), report opposite findings. Why the conflict in findings?

A look at Brady's assignment procedure suggests one possible reason. Only four pairs of monkeys were used, and they were assigned to experimental and yoked control conditions on the basis of pretest avoidance scores. The monkeys assigned to the executive condition were those that had the highest rate of avoidance responding. The other four monkeys became the yoked control participants. Unfortunately, rate of avoidance responding and ulceration are positively related. Therefore, those assigned to the executive condition were probably more prone to ulceration than were the controls. Because of this, it is difficult to conclude that having control over aversive environmental events (executives) leads to greater physiological debilitation than not having control (employees). Indeed, the research of Weiss, who used random assignment and the yoked control procedure, supports the opposite conclusion.

Another example of the use of yoked control can be seen in a long series of studies in rats that attempt to understand the detrimental effects of sleep deprivation. Note that the word *detrimental* is appropriate because rats that experience continual and total sleep deprivation die in about 16 days. As you might imagine, the basic design includes a group of rats that are deprived of sleep (experimental group) and a

group of rats that are not deprived of sleep (control group). A common method to deprive rats of sleep is to force them to move (using some form of treadmill) when the first sign of sleep appears in their recordings of brain activity. Do you see the potential confound? Not only will the experimental rats be deprived of sleep, they will also engage in much more physical activity. Thus, any observed effects could be due to sleep deprivation or could be due to physical activity.

To eliminate this potential confound, Allen Rechtschaffen, a sleep researcher, devised an experimental methodology that incorporated a yoked control. The experimental apparatus consists of a Plexiglas box with a pool of water at the bottom. Elevated above the pool is a circular disc that can rotate and a wall that separates one half of the box from the other. Two rats are placed on the circular disc, one on each side of the wall. Whenever the experimental rat begins to fall asleep, the circular disc begins to rotate. To avoid hitting the wall and falling into the pool of water (which rats do not like), the experimental rat will awaken and begin walking in the opposite direction. Because the control rat is on the same circular disc, the control rat must engage in the same amount of walking behavior. However, the control rat can sleep anytime that the experimental rat is awake. Thus, the experimental rat is deprived of sleep while the control rat is not (at least to a great extent), and the amount of physical activity is controlled. Even with this elegant design, there are several other potential confounds that the researchers needed to control. Can you identify some of them?

Placebo Control Procedures. The word *placebo* comes from the Latin verb meaning "to please." The evolution of the term from both a research and therapeutic perspective is interesting. Medical historians have noted that almost any kind of treatment used in the early days of medicine seemed to have therapeutic properties. Even though these therapeutic treatments had no obvious direct relationship to the problem being treated, they did, in fact, alleviate distress. Remedies such as lizard blood, bat blood, crocodile dung, frog sperm, putrid meat, hoof of ass, and others were used (Shapiro & Morris, 1978). Often, complex rituals were used, and the ingredients that were given caused bodily discomfort. These rituals and the administration of the "therapy" had the effect of arousing faith and also the expectation that the "therapy" would be effective. Rituals of a different kind are still used by physicians, therapists, and researchers.

Today, physicians and therapists are well aware of the placebo effect and often use it to their advantage. The phenomenon is well documented in medicine and psychotherapy. Interest in understanding more about the placebo is very high, and prominent medical and physiological investigators are giving serious attention to studying it. There is now considerable evidence indicating that the placebo can act like medication and can result in marked physiological changes. Placebos have been shown to actually alter the body chemistry and to mobilize the body's defenses. There is also evidence showing that administering a placebo has an effect on the neurochemistry of the brain. In studies dealing with the experience of pain, placebos have been shown to significantly reduce pain. These studies

have shown that placebos trigger the brain to release internal opiates (endorphins), which are known to have a marked effect on the experience of pain. Why and how placebos do this is still a mystery.

The placebo effect is similar to demand characteristics, in that cues or treatment in the situation give rise to expectations on the part of the participant. However, demand characteristics are more idiosyncratic and vary among individual participants, whereas the placebo effect tends to be specifically and directly tied to the treatment condition. The placebo effect in research settings is generally seen in outcome studies where different drug treatments or therapies are being compared. Outcome studies of treatments are different from experimental studies where other kinds of interests are evaluated. In outcome studies comparing different therapies or different drugs, a **placebo control** group is necessary simply to assess the therapeutic effects of believing that one has received a curative treatment.

We can illustrate the need for this type of control. For example, if we were interested in evaluating the effectiveness of a therapeutic drug, we would have to untangle the actual effects of the drug from the expectation that the drug has a therapeutic effect. Simply believing that a drug or a therapeutic treatment has an effect can lead to a consistent and marked change in behavior. In our experiment, we would have two groups. One group would be responding to the drug *and* to whatever placebo effect it might have; the other group would respond to the placebo effect alone. In the latter case, the placebo would be an inert substance or sugar pill that appeared exactly like the drug itself. If differences in behavior followed, then we would attribute these differences to the effects of the drug because we have controlled for the placebo effect by allowing it to occur in both groups. The placebo is used to ensure that participants in both groups have the same expectations and beliefs. As noted, for the placebo to be effective, it must be indistinguishable in appearance from the actual drug. In some cases, it may be necessary to provide side effects similar to those experienced with the actual drug. When using a placebo control procedure, the expectations of participants are distributed equally among the groups used in the experiments.

The placebo control group method for evaluating the effects of the independent variable is a powerful technique, but under some circumstances, its use presents real problems. Both ethical and practical problems can arise when evaluating the effectiveness of clinical-therapeutic procedures, such as psychological or medical treatment. Several recent articles have addressed the use of placebo controls in the development of new drugs to treat Alzheimer's disease (Kawas, Clark, & Farlow, 1999) and schizophrenia (Fleishhacker & Marksteiner, 2000). When evaluating a therapeutic technique, the placebo control method requires that a sample of individuals suffering from an illness or disorder be divided into at least two groups—one that receives the treatment therapy and one that receives the placebo condition. For example, in the case of a drug therapy, the evaluation requires that the drug be given to the treatment group and be withheld from the placebo group—even though the drug might in fact help them. What do you do under these circumstances when, partway through the study, the drug appears to be effective for the treatment group? This becomes an ethical issue. How do you maintain patients for the placebo control

condition after information becomes available suggesting that a treatment is effective? This poses a practical problem.

Some alternatives to the placebo control method are available, in addition to the crossover design. One is to use an **active control** procedure. With this procedure, half the patients receive an established treatment whose degree of effectiveness is known; the other half receives the new treatment. Another method is the **historical control** procedure. In this case, the new therapy is compared with clinical records of past patients who were untreated or who received another therapy. In still other instances, if a drug or treatment clearly reduces pain, or lowers blood pressure, or prevents suicide, or *dramatically* reduces any important physical or behavioral problem *consistently*, then a placebo control study may not be necessary. There are also instances in which all participants in a research study receive the treatment but in varying doses. Sometimes, therapies have been adopted on the basis of the therapists' observations rather than controlled studies. Although controlled studies are far better for evaluating a therapy, it is sometimes difficult to argue with the observation that "people used to die without a drug or treatment but now live with it."

Case Analysis

Researchers continue to investigate the process by which children learn to read. Debate continues over the most effective method to teach reading. The whole-word method, used in a majority of schools, emphasizes the reading of literature and understanding words by understanding the context. The phonics method emphasizes the decoding of words by focusing on the sounds of the alphabet. A researcher decides to compare two classes that will use the two different methods.

The researcher knows that School A uses the whole-word method and that School B uses the phonics method. The researcher randomly samples one first-grade class from School A and one first-grade class from School B. At the end of the school year, the researcher receives consent to administer a standardized reading test to the children in both classes. Scores show that children in the phonics class scored significantly higher on the reading test than children in the whole-word class. The researcher concludes that the phonics method is more effective than the whole-word method.

Critical Thinking Questions

1. Do you believe that this conclusion is warranted?
2. Does this represent a true experimental design? Why or why not?
3. Which extraneous variables were effectively controlled by the methodology used?
4. Which extraneous variables were not effectively controlled by the methodology used?
5. Which control techniques would be effective in reducing the extraneous variables?
6. Which experimental design would be more effective, and how would it be conducted?

General Summary

A true experiment is characterized by a high degree of experimental control, the hallmark of which is random assignment. The research techniques described in this chapter provide the tools with which a researcher can confidently answer questions in the field of behavioral research. The challenge is to understand the tools that are available and to know when to use them. The good researcher knows that extraneous variables can creep in at every step of the research process. Researchers can control extraneous variables through the experimental setting, consent, instructions, sampling techniques, assignment techniques, observation techniques, measurement techniques, interactions with participants, and the use of research designs with control groups.

The experimental setting should be selected so that the researcher can control the stimuli and events that the participant will experience, without sacrificing external validity. Informed consent and experiment instructions should be carefully worded to avoid demand characteristics, evaluation apprehension, diffusion of treatment, and other unwanted effects. In addition to random sampling of participants, random assignment to groups can be a very effective deterrent to extraneous variables. In other situations, random error due to individual differences or systematic error due to a confound can be controlled by pairing the scores in the groups via natural pairs, matched pairs, or repeated measures designs. Precise observation and measurement techniques are critical, along with a professional demeanor on the part of the experimenter. Finally, use of a control condition that does not receive the treatment provides a critical comparison condition against which to judge the effect of the treatment in the experimental condition. A control condition is particularly important when placebo effects are a concern.

Before we turn to specific experimental designs, the next chapter will review some basic statistical issues.

Detailed Summary

1. A variety of research design techniques are available to control extraneous variables. Random assignment and the extensive use of other control techniques are hallmarks of the true experiment.
2. The true experiment has advantages and disadvantages. Advantages relate to the extensive use of control to reduce the influence of extraneous variables and thus alternative interpretations of the data. Disadvantages, in some cases, relate to the ability to generalize findings to the “real” world. In addition, some research questions cannot be addressed with a true experiment, either because the independent variable cannot be manipulated or because it would be unethical to do so.

3. Control techniques include control through sampling, assignment to conditions, setting, consent, instructions, observation, measurement, experimenter interactions, use of control groups, and research design.
4. Control through sampling refers to the use of random sampling to increase both internal validity (reduce systematic and random error) and external validity. Sampling refers to the way in which participants are selected for the study. With random sampling, each element in the population has an equal chance of being in the study. When random samples are not practically possible (you then have a sample in search of a population), the researcher should nevertheless strive for the characteristics of random sampling.
5. Control through assignment to conditions refers to the ways in which research participants who have been sampled for the study are assigned to particular groups or conditions. Control can be exercised by the use of random assignment to create an independent samples design or the use of correlated samples designs (natural pairs, matched pairs, repeated measures) in which control is achieved by pairing scores in the different groups/conditions.
6. Control through experimental setting refers to the researcher's ability to determine stimuli and events in the research setting. Greater control of the setting reduces sources of systematic and random error.
7. Control through consent refers to the use of relaxed but professional language during informed consent. We do not want to create anxiety, and we do not want to reveal details that might lead to unnecessary demand characteristics. In many situations, we can use a single-blind technique in which participants are not aware of the group or condition that they are in.
8. Control through instructions refers to the use of identical instructions for all participants (unless the instructions serve as the IV). It often includes a request to not discuss the study with other individuals who might participate in the future.
9. Control through observation and measurement refers to observation conducted without participant awareness and/or the use of a double-blind technique, whereby participants do not know what group they are in and observers do not know what group participants are in when observations are made. Clear operational definitions and periodic checks of recording equipment also contribute to control.
10. Control through experimenter interactions refers to the professionalism with which the experimenter presents him/herself.
11. Control through use of control groups and research design refers to research designs that include a comparison group that receives either a different treatment or no treatment (control group). Examples of such designs are the randomized pre- and posttest control-group design and the randomized posttest-only control-group design.
12. Specific types of control groups include the yoked control and placebo control. Yoked control refers to an experimental situation in which two participants experience exactly the same environmental

events at the same moments in time with the exception of the one variable that is manipulated by the researcher (the IV). Placebo control refers to the fact that for some treatments (especially drug treatments), participants readily expect specific effects, and these expectations can mimic the effect of the treatment. A placebo control group provides a measure of these expectancy effects when the treatment is not actually administered.

Key Terms

active control
correlated samples
double-blind study
historical control
independent samples
matched pairs
natural pairs
placebo control
random assignment
repeated measures
single-blind study
yoked control

Review Questions/Exercises

1. Briefly describe a hypothetical study of emotions in which random assignment of participants to treatment conditions is possible. Briefly describe a hypothetical study of emotions in which random assignment is not possible. Briefly describe a hypothetical study of emotions in which random assignment is unethical.
2. Consider a research topic in which you are interested, and conduct a database search for journal articles on that topic using a database available on your library website. Locate a study that is a true experiment, and provide a summary of the experiment.
3. Read the Method section of the article you chose in question 2. Using the summary of control techniques illustrated in Figure 9.1, identify which of the techniques were used in the experiment, and provide a description of how each was used.