In the previous chapter, we discussed the concept of validity and the various threats to the validity of research. In this chapter, we talk about ways of counteracting those threats to the validity of psychological research. First, we define the concept of control. Then we discuss, in turn, several general strategies for achieving control. Finally, we describe a number of specific control strategies.

**THE CONCEPT OF CONTROL**

**Control** is the other side of the validity coin. The heart of the experimental approach to knowledge is to ask the following two questions: (1) What are the threats to the validity of a contemplated piece of research? (2) What means are available to neutralize those threats? This approach is so basic that anyone who is acquainted with research has heard of control groups. Every experiment must have a control group. Right? Wrong.

It *is* true that you must have some method of countering every plausible alternative explanation of the results of your experiment. It is also true that this often involves the use of a group of subjects who do not experience the manipulation—that is, a control group—to serve as a standard against which to compare the effect of the variable of interest. Many experiments are performed, however, in which the use of a group that does not receive the independent variable makes no sense at all. Suppose you are interested in the
effect of teaching methods on learning. You might arrange for some students to receive only lectures and others to receive only discussion. You conclude that one method of teaching was better than the other without having a group that never went to class.

We will define control as any means used to rule out possible threats to the validity of a piece of research. In psychology, the concept of control is used in two ways. The fundamental meaning of the term is that of providing a standard against which to compare the effect of a particular independent variable. If two experimental conditions differ on only one independent variable, then any difference between the two conditions following the treatment may be attributed to the operation of that variable. All other explanations are ruled out by the existence of the second, or control, condition.

Recall our discussion of Mill’s methods for determining causes in Chapter 2. Following the method of difference, if two individuals differ on only one variable, then that variable may be considered to be the cause of the difference between the individuals. The concept of control is essentially a way of establishing that two individuals, or groups, or conditions, are identical except for the variable of interest. When that is the case, then the research is valid, and the method of difference can work.

This meaning of the term control is illustrated in Table 6.1. Two groups of subjects are tested on a dependent variable. Group 1 receives Treatment A; it is the experimental group. Group 2 receives no treatment; it is the control group. The control group serves as the basis of comparison for the experimental group. If the two groups were equal before the experimental treatment, then any postexperiment difference between them can be attributed to the treatment.

Although a control group is an effective way of achieving control of extraneous variables, it is not the only way. We said earlier that control can be achieved without a control group. Table 6.2 illustrates this point. We still have two groups, both of which are tested after receiving treatment. However, instead of Group 1 experiencing A and Group 2 experiencing the absence of A, both groups experience some value of A. As suggested previously, A₁ and A₂ could be two different teaching methods. Assuming the groups were equal before treatment, we can attribute any difference between Group 1 and Group 2 on the test to the difference between Condition A₁ and Condition A₂.

<table>
<thead>
<tr>
<th>Table 6.1 Use of a Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Group 1 (experimental group)</td>
</tr>
<tr>
<td>Group 2 (control group)</td>
</tr>
</tbody>
</table>
control condition
a condition in an experiment that does not contain the experimental manipulation
within-subjects experiment
research design in which each subject experiences every condition of the experiment
between-subjects experiment
research design in which each subject experiences only one of the conditions in the experiment

Although we do not have a control group as such, each group serves as a control for the other. We have as much control in this situation as we did in the previous example, in which we had a control group.

Let us consider one further point. Instead of having different subjects experience each condition, in some experiments each subject experiences every condition. In such an experiment, instead of having a control group, we have a control condition, as illustrated in Table 6.3. When each subject experiences every condition, we say that each subject serves as his or her own control. An experiment of this kind is called a within-subjects experiment because the differences between conditions are tested within individual subjects. An experiment in which different groups of subjects experience different conditions is a between-subjects experiment because the differences between conditions are tested between different subjects.

A second meaning of the term control is distinct from the first but closely related—namely, the ability to restrain or guide sources of variability in research (Boring, 1954, 1969). This idea of experimental control is the one brought home so convincingly by the operant conditioning work of B. F. Skinner. When one has so limited the sources of variability in an experiment that the behavior becomes highly predictable, one has achieved experimental control. We are extremely impressed to observe a pigeon that has been trained to peck a key for food in the presence of a green light but not to peck in the presence of red. When the bird is well trained, the light virtually turns the bird on and off.
The two meanings of control are related in the following way. The primary meaning allows one to conclude that a dependent variable is associated with an independent variable and not with any other variable. The second usage facilitates drawing this conclusion by so limiting the number of variables operating in the situation and their range of values that the conclusion is clearer.

We can characterize the difference between the two meanings by use of the terms control experiment and experimental control (compare Sidman, 1960). When we have experimental control (secondary meaning), we have a much more sensitive situation in which to rule out alternative explanations of the experimental results (primary meaning).

Both meanings of the term control relate to the use of statistics in research. First, we use inferential statistics to evaluate the probability that a difference between experimental and control groups or conditions is likely to have arisen by chance alone. Second, we make enough observations or use enough subjects to reduce the variability of our estimate of the size of the experimental effect, and so make our statistical evaluation more precise.

The next section deals with the most important ways of achieving control in research. At the outset we should note that just as all types of threats to validity do not appear in all research, so it is not necessary or even possible to use every means of control in all research. The various methods of control are tools for psychologists to employ as necessary. Some will be used almost always, others less often.

GENERAL STRATEGIES

We will discuss three general strategies for achieving control in psychological research: using a laboratory setting, considering the research setting as a preparation, and instrumenting the response. Although these strategies are closely related, we will consider them separately for emphasis.

Control in the Laboratory

Laboratory research is generally preferred to field research. The reason is simple and has to do with what a laboratory is. We tend to think of a laboratory as a room with gray or black furniture, no curtains on the windows, tile floors, and workers dressed in white coats. Certainly, we are describing a typical laboratory, but the description has nothing to do with the essentials. Basically, a scientific laboratory is a place set up to allow the most appropriate control over variables of interest in the particular research. Thus, a social-psychology laboratory might well have rugs on the floor, curtains on the windows, pictures on the walls, and comfortable chairs—like any living room. Laboratory work in social psychology requires control over elements such as choice of subjects, beginning and end of social interaction, and freedom from distraction. If someone's home or a storefront building meets
these requirements, such a setting might serve just as well or better than a sterile-looking room. The results of laboratory research depend entirely on the degree and type of control that is possible.

In Chapter 7 we will discuss some methods and advantages of field research. Here we will simply say that at times field research is preferable and at times laboratory research is preferable. Much social research is done in field settings because it is not possible or ethical to manipulate certain variables. Most people would frown on mugging subjects to learn what determines whether they will call the police. In some cases, the effect of a manipulation may not be realistic enough in the laboratory. Simulating a riot, for instance, would be difficult to do in a laboratory. But even those people who advocate field research agree that they must give up a degree of control and that problems of validity thus become greater. Field research is warranted when ethical or practical problems preclude the degree of control that would justify calling a certain research program a laboratory experiment. Laboratory research remains the ideal simply because the maximum control consistent with the nature of the problem is the ideal.

The Research Setting as a Preparation

One of the first questions to be answered in designing research is to decide what type of setting you want to use. You may be interested in learning what determines whether people will be cooperative or competitive. You could study how children play with toys, how basketball players pass to each other, or how salespeople decide who will help a customer. Any of these might be a good study. Another way to do research on cooperation is to use a situation called the prisoner’s dilemma. The name comes from the situation in which two people have been arrested for involvement in a crime. If both of them refuse to confess, both will get a modest punishment. If both confess, they both get an intermediate punishment. If one confesses and the other does not, the one who confesses will be pardoned (rewarded) and the other severely punished. The dilemma arises because confession can lead either to reward or to punishment, with nonconfession also leading to punishment. This situation has been adapted to the laboratory, and much research has been done on the effects of the magnitude of rewards and punishments, relationship of the two people involved, and so forth.

The idea of a preparation is familiar to anyone who has studied biology. Researchers often use the giant nerve axon of the squid to study nerve conduction. Because the squid nerve is much larger than those in other animals, it permits biologists to do things that they cannot do with other nerves. The concept of a preparation is not as familiar in psychology, yet one of the researcher’s goals is to choose the most suitable preparation for studying a given problem. Some of the most important contributions to psychology have been made by people who devised a new preparation for studying a given phenomenon. Perhaps the best example is that of B. F.
Skinner, who created the device that everyone, except Skinner, calls the Skinner box.

Before Skinner’s invention, a number of devices had been used to study learning in small animals. Such study began with a maze patterned after the Hampton Court maze in England, one of those people-sized hedge mazes that were popular on large estates a few hundred years ago. Researchers soon realized that this first maze was too complicated for its purpose, so mazes were made progressively simpler until the T maze was designed. Later someone made a runway with no choice points at all and measured the speed at which rats ran the “maze.” Then Skinner took a box, added a lever, and the rest is history. What his apparatus provided that others did not was simplicity, plus the opportunity for the rat to respond as often as it liked without waiting for someone to pick it up and return it to a starting point.

Skinner’s seemingly small change made it possible to study rate of responding rather than number of correct turns or even speed in the runway. Having focused on an important dependent variable largely overlooked by others, Skinner went on to revolutionize the study of learning in animals. Thus, the Skinner box became one of the most important experiment preparations in psychology.

Preparation in this sense is involved in every experiment, but hardly ever is it spoken of that way in books on research design in psychology. Nevertheless, it is one of the most important considerations in designing research. Exactly which situation will provide the most powerful relationship between the variables of interest? No amount of sophisticated design or statistical analysis will make up for a poor choice of research preparation.

### Instrumentation of the Response as Control

We have discussed the research setting as a preparation that allows the sensitive analysis of a phenomenon. Another important means of increasing the sensitivity of the research is to improve the measurement of the behavior being studied. Many researchers pay little attention to the response that will be measured, but here is where a little effort can pay big dividends. Just as certain preparations have become classic in psychology, so certain methods of measuring dependent variables have also had enormous impact. We have already mentioned the Taylor Manifest Anxiety Scale. The Minnesota Multiphasic Personality Inventory, the polygraph, and Stevens’s direct psychophysical scaling methods are other excellent examples. These techniques have greatly improved the sensitivity of research in their respective fields. Many scientists devote their careers to developing and honing measuring devices. Most of these people do not intend to become methodologists but do so in order to evaluate more precisely phenomena of interest to them.

I have used the term instrumentation in discussing the task of improving response measurement. This use may seem strange because not all measurement methods employ mechanical means. The usage is deliberate, though. It calls attention to measuring devices as instruments for reducing
behavior to numbers or to other forms convenient for data analysis. A characteristic of a good measuring instrument is that it takes the response out of the realm of casual observation and makes it reliable. Only in this way can we speak of measurement of behavior as objective, thus meeting the requirement of interobserver reliability necessary for science. Therefore, even a measure of a subjective state, such as the pleasantness of an odor, can be considered objective, provided the instrumentation of the response is adequate.

**SPECIFIC STRATEGIES**

- **Subject as Own Control**

We are aware that each of us is unique and varies in many ways that could be important in an experiment. One of the most powerful control techniques is to have each subject experience every condition of the experiment. In this way, variation caused by differences between subjects is greatly reduced. The experimenter is wise to adopt the strategy of using subjects as their own controls whenever possible.

This control method is common in many areas of psychology, particularly sensation and perception. For example, if you are interested in the effect of adaptation to different concentrations of salt on the threshold for salt, using different subjects for each condition does not make much sense. The experimental manipulation is not likely to destroy the naivety of the subject, because the subject is unlikely to guess the purpose of the experiment even after experiencing it. In fact, the subject may not be aware that the experiment has different conditions. In addition, if enough time is allowed between conditions, there is not likely to be an important carryover between conditions. The subject will recover in a few minutes from the effect of adaptation to salt and will be ready to experience the next condition.

In many experiments, however, using subjects as their own controls simply is not possible. Once the subject has learned a problem by one method, learning the same problem again using a different method is impossible.

Another situation in which using subjects as their own controls is not feasible occurs when contrast effects exist between the conditions of the experiment, so that experiencing one condition may influence the response to another condition. These contrast effects, also known as order and sequence effects, will be discussed in Chapter 9. For now it is sufficient to note that there are situations in which conditions may affect one another. For instance, if magnitude of reward is the independent variable, subjects who experience a large reward first may respond less to a small reward than they would if only the small-reward condition had been received.

A more concrete example is provided by an experiment involving lifted weights. Suppose that you have two weights and each is lifted only once. You
may think that you have controlled for order and sequence effects if half of the subjects experience one order (light, heavy) and the other half the opposite order (heavy, light). Let us suppose that there is a "true" response of 6 to the light stimulus and 8 to the heavy one, as determined in a between-subjects experiment in which each subject lifted only one weight. Now, if there were a contrast effect between the two weights that resulted in a doubling of the true difference in the responses when the weights were presented sequentially, the subjects who experienced light before heavy would give responses of 6 and 10, instead of 6 and 8. Those who experienced heavy before light would give responses of 8 and 4. The average responses for the two different orders would thus be 5 and 9, instead of 6 and 8, and the difference between the stimuli would appear to be 4, instead of the true difference of 2.

The type of effect we have just described is summarized in Table 6.4. Sometimes these effects can simply exaggerate an outcome that would occur between subjects, as in this example. Other times they produce outcomes that would not otherwise be found. The difference between using a within-subjects design and a between-subjects design can cause puzzling discrepancies in the results of experiments.

Returning to the lifted-weights example, we note that Harry Helson and his students (Helson, 1964), in their study of sequence effects in perceptual judgments, gave us the area known as adaptation-level theory. They found that a given weight would be called light or heavy depending on what other weights were presented along with it. Effects such as these can be studied only by using designs that permit contrast effects.

<table>
<thead>
<tr>
<th>TABLE 6.4 THE PROBLEM OF CONTROLLING FOR SEQUENCE EFFECTS IN WITHIN-SUBJECTS EXPERIMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WITHIN SUBJECTS</td>
</tr>
<tr>
<td>ORDER</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Light, heavy</td>
</tr>
<tr>
<td>Heavy, light</td>
</tr>
<tr>
<td>Average effect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BETWEEN SUBJECTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>STIMULUS</td>
</tr>
<tr>
<td>LIGHT</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>&quot;True&quot; effect</td>
</tr>
</tbody>
</table>
In summary, you should consider using subjects as their own controls whenever three conditions can be met: (1) Using subjects as their own controls is logically possible. (2) Participating in all conditions of the experiment will not destroy the naivete of the subject. (3) Serious contrast effects between conditions will not be present.

**Random Assignment**

Another powerful control method is **random assignment** of subjects to conditions. The term *random assignment*, or *random allocation*, is used here in a specific sense: The allocation of subjects to conditions is random when each subject has an equal and independent chance of being assigned to every condition. The advantage of random allocation of subjects to conditions is that once subjects have been randomly assigned, the only way that confounding of subject-related variables with the experimental variable can occur is by chance.

At first glance, calling random allocation a method of control may seem improper because you appear to be throwing away a means of control and casting yourself on the mercies of chance. However, when you randomly assign subjects to conditions, you can be sure that only chance could cause the groups to be unequal with respect to a potential confounding variable. All sources of confounding variables are ruled out except as they become associated with the conditions by chance. Even if some variable is confounded with the independent variable by chance alone, assessing the likelihood of this happening is possible via statistical methods.

Modern statistical analysis provides numerous ways of testing whether experimental results are likely to have occurred by chance alone. These statistical methods rely completely on random allocation of subjects to conditions. In other words, statistical tests make estimates of the probability that purely random allocation of subjects to the various experiment conditions might have produced the results obtained. If the subjects are not assigned to conditions randomly, the statistical tests are not valid. One of the biggest “sins” in experimental psychology is to perform an experiment that cannot be analyzed statistically. The surest way to commit this error is to fail to assign subjects to conditions randomly.

Suppose you have already chosen your subjects from a population in some manner, and you have an experiment with two conditions. How do you decide which subjects will experience which conditions? There are many different ways of accomplishing this, but the following is a good way to do it.

The first step is to assign numbers to individuals. If you have 20 subjects, you assign numbers 1 through 20 to the individuals. Then go to a random-number table, such as the one in Table 8.1 (page 206). You could decide that the first subject would go into Group 1 or Group 2 depending on whether a 1 or a 2 came up first in the random-number table. This, however, would require you to go through many numbers, because only about one-fifth of the numbers in the table are 1s and 2s. (Of course, if you had more than
two groups, it would be most efficient to consider the numbers in the table to be the number of the group.

An alternative would be to consider Group 1 to be odd and Group 2 even. You would then put a subject into Group 1 if an odd number came up and into Group 2 if an even number came up. Turn to the random-number table on page 206 (Table 8.1). Because the first two columns appear to have been used, we might as well take column 3. The first four digits are all even, so the first 4 subjects will go into Group 2. Then there is one odd number followed by two more even numbers, one odd number, and then four more even numbers. By this time, we have allocated 12 subjects to conditions: 2 into Group 1, and 10 into Group 2. Because we want 10 per group, the last 8 must all go into Group 1. Does this sound like a random assignment—to have 8 out of the first 10 subjects go into Group 2 and all of the last 8 go into Group 1? It may not look random, but it was the result of a random process. Some researchers would be tempted to throw out this particular result as not representative and start over again, but the statistical model says we should go with it. (I would probably do another random assignment if I had some reason to believe that the numbers assigned to the subjects were systematic in any way—for example, if the assignment seemed to follow the order in which the subjects had signed up for the experiment.)

Using the procedure of randomization requires care. Students may be asked to volunteer for an experiment by signing up on a sheet for available times. The experimenter might be tempted to assume that the times are selected randomly and might place the first half of the students in one condition and the last half in the other. You can easily think of a number of ways in which the two groups could differ. The people who sign up for the early times may be more highly motivated to serve as subjects. On the other hand, all may have jobs requiring them to leave campus in the afternoon. The experimenter must randomly assign subjects to conditions after they have signed up.

Another example of a mistake in allocating subjects is to pull rats from a group cage and place the first batch selected in Condition A, the second batch in Condition B, and so forth. A little reflection will make it clear that the order in which the rats will be picked depends on their tendency to approach or avoid the experimenter’s hand. This difference could be relevant for many experiments in learning, motivation, or social behavior. Table 6.5 summarizes the steps in a randomized-groups experiment.

---

**Matching**

Experimental precision can sometimes be improved by **matching** subjects on a pretest before randomly allocating them to conditions. When the subjects differ among themselves on an independent variable known or suspected to affect the dependent variable of interest, matching may be necessary. For example, suppose you are studying the effect of two different feeding schedules on weight gained by rats. You might expect that rats that were
TABLE 6.5 Steps in conducting a randomized-groups experiment

1. Randomly assign subjects to groups.

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁</td>
<td>S₂</td>
</tr>
<tr>
<td>S₃</td>
<td></td>
</tr>
<tr>
<td>S₄</td>
<td></td>
</tr>
<tr>
<td>S₅</td>
<td>S₆</td>
</tr>
<tr>
<td>S₇</td>
<td></td>
</tr>
<tr>
<td>S₈</td>
<td></td>
</tr>
<tr>
<td>S₉</td>
<td>S₁₀</td>
</tr>
</tbody>
</table>

2. Administer experimental conditions.
3. Examine differences between groups.

Heavier to begin with would continue to gain more weight regardless of the schedule they were on. If subjects were allocated randomly to conditions, more of the heavier rats could wind up in one condition than in the other. By weighing the rats before the experiment, you can allocate them in such a way that the average weight in the two groups is the same.

The first requirement to justify matching is a strong suspicion that there is an important variable on which the subjects differ that can be controlled by matching. Further, you must believe that a substantial correlation will be present between the matching variable and the dependent variable. In our example of the weight-gain experiment with rats, you would find the two lightest rats and randomly place one in Group A and the other in Group B. You would repeat this procedure until you had paired off all the rats. If you found that, in fact, those animals that were initially heavier tended to gain weight regardless of the group they were in, you would have been justified in matching on weight. By correlating weight gain with starting weight, you would have found that weight gain correlated with beginning weight in spite of the effect of the variable of interest. On the other hand, if you found that there was little or no correlation, you would have wasted your effort in matching the subjects.

In fact, you can weaken your experiment by matching the subjects if the matching variable is not substantially correlated with the dependent variable. This effect results from the fact that the statistical test appropriate for a
matched-groups design considers the data from pairs of subjects, while the randomized-groups test considers individual subjects. You can see that there are twice as many subjects as pairs of subjects, so the randomized test has more numbers to work with and therefore is more powerful.

A second condition necessary to justify matching is that it must be feasible to present a pretest to the subjects before assigning them to the conditions. For example, weighing rats before an experiment would be a simple matter, but giving an IQ test to every prospective student in an experiment on learning may not be feasible. The experimenter’s time would probably be better used in simply testing more subjects, unless the IQ data can be obtained readily and ethically or unless the experiment is long enough to allow time for IQ testing.

Some bases for matching are better than others. Generally you try to match on the basis of some variable that has the highest possible correlation with the dependent variable. Normally the highest correlation is between the dependent variable and itself. In other words, if you are doing an experiment using reaction time, matching subjects according to their reaction times makes the most sense. You could present some practice trials and then allocate subjects to conditions based on their performance in the trials. On the other hand, if you are doing an experiment on learning and you want to control for intelligence, matching according to socioeconomic status (SES) would be a poor choice, even though there is some correlation between SES and intelligence scores. Your choice would be poor because the correlation between SES and IQ is weak. The slight control achieved by matching would be offset by the lower statistical power of a matched-groups design.

Let us emphasize a final point about the mechanics of matching. Even when you have matched your subjects, you must still randomly allocate the members of the pairs to conditions. If you have ten pairs of rats matched for weight, you must flip a coin or follow some procedure that will ensure that the members of each pair are allocated to groups randomly. Otherwise, your procedure for placing them into groups could introduce confounding. Table 6.6 summarizes the steps in a matched-groups experiment.

---

**Building Nuisance Variables into the Experiment**

Another way to handle variables that cannot easily be removed from the experiment is to design the experiment so that these nuisance variables become independent variables in the study.

Suppose that your subject pool consists of both day-school and night-school students in introductory psychology. These people may differ in several ways that could relate to psychological variables. Night students may be older, may have more family and work responsibilities, and so forth. If you suspect that your subjects are dissimilar on some dimension related to day-versus night-student status, you have two choices. The first is to use only day or night students. This solution has the advantage of reducing the variability, but it also reduces the subject pool and the generality of the results. The second choice is to build a nuisance variable into the experiment.
**TABLE 6.6 Steps in Conducting a Matched-Groups Experiment**

1. Administer pretest.
2. Rank subjects on pretest.
   
   \[
   S_1 \\
   S_2 \\
   . \\
   . \\
   . \\
   S_{10}
   \]

3. Form pairs on the basis of rankings.
   
   \[
   \begin{align*}
   &S_1 \quad \{ \\
   &S_2 \quad \{ \quad 1st \text{ pair} \\
   &S_3 \quad \{ \quad 2nd \text{ pair} \\
   &. \quad . \\
   &. \quad . \\
   &. \quad . \\
   &. \quad . \\
   &S_9 \quad \} \quad 5th \text{ pair} \\
   &S_{10} \quad \} \\
   \end{align*}
   \]

4. Randomly assign members of pairs to groups.
   
   \[
   \begin{array}{c|cc|c}
   \text{GROUP A} & \text{GROUP B} \\
   \hline
   S_1 & S_2 \\
   S_4 & S_3 \\
   S_5 & S_6 \\
   . & . \\
   . & . \\
   . & . \\
   S_{10} & S_9 \\
   \end{array}
   \]

5. Administer experimental treatments.
6. Examine differences between members of pairs.
   
   \[
   S_1 \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quarter

What is the average difference between pairs?
Figure 6.1 shows the results of a hypothetical experiment in which night- or day-student status was designed into the experiment as a nuisance variable. If we do not consider day or night status, we find considerable overlap between the results of Conditions A and B. We might not be willing to conclude that the conditions had a differential effect on the dependent variable. (We are ignoring the possibility of using inferential statistics to help in this decision.) But let us consider day or night classes as a nuisance variable; that is, let us analyze the data separately for day and night students. We find now that no overlap exists between Group A and Group B for night-school students considered alone or for day students considered alone. We have increased the sensitivity of the experiment by building a nuisance variable into the study. Note that the nuisance variable need not have any theoretically important role in the experiment. We do not care why night-school students scored higher. On the other hand, the nuisance variable may suggest new theoretical questions for another experiment.

### Statistical Control

Except for randomization, all of the control methods described so far can be classified as methods of achieving experimental control; they aim to reduce variability as much as possible. Sometimes these techniques can be spectacularly effective and a very few observations on a single subject can be used to draw firm conclusions. Usually, however, that old devil variability cannot be completely exorcised from the experiment. Then it is necessary to use **statistical control**.

Statistical control in the broad sense is synonymous with inferential statistics, the branch of statistics that deals with making decisions in the face of uncertainty. Suppose you have a difference between two groups on some dependent variable in an experiment. Was the effect real, or did it happen by
chance? The point behind the question is that statistical control is involved in designing an experiment. Are there enough subjects? How many trials should there be? Can the experiment as designed be analyzed properly by accepted statistical methods?

Such considerations are important enough to merit courses that specialize in the statistical analysis of experimental data. For now, because you are not taking such an experimental-design course, you will have to answer questions of statistical control with general knowledge about statistics and with advice from your instructor. Remember, too, that in the end the question of statistical control comes down to whether you and others believe the data. If you have enough subjects to look convincing, and if you have avoided the pitfalls discussed in this chapter, you probably have a good experiment. Although you may be fooled into thinking that some effect was caused by the independent variable when it was actually random—especially when you look at your own data—at this stage of your career you should aim to design an experiment that will convince yourself and others.

In the narrower sense, statistical control refers to a means of equating subjects on paper when they cannot be equated in fact. Suppose you are studying the effect on grades of two different teaching methods in the classroom. Because randomly constituting the classes is not feasible, you must work with existing classes. If the students in the two classes do not have the same average IQ, you will have a problem in attributing the difference in grades to the teaching methods. But if you know the relationship between IQ and grades in the class, you can find out how much the students' grades differed from what you would have expected, based on the prediction of the IQ/learning relationship.

This approach is illustrated in Figure 6.2, which represents the relationship between grades and IQ. Each data point represents a single subject, the X being placed at the intersection of the person's grade

![Figure 6.2 Using statistical control to compare subjects.](image-url)
(measured on the y-axis) and his or her IQ (on the x-axis). The slanting line shows what grade would be predicted for persons having particular IQs. (In other words, it is the regression line predicting grade from IQ.) The data point circled and labeled A represents a single subject's position on the two axes. You can see that Subject A earned a higher grade than would have been expected from the relationship shown by the slanted line. Subject B, however, earned a lower grade than predicted by the line. Notice that both students received the same grade.

The basic idea of statistical control is that it enables you to compare students not on their absolute grade, but on the difference between grade and what would have been expected from the line predicting grade from IQ. Subject A would be scored as earning a grade of plus-so-many points and Subject B as earning minus-so-many points. You could conclude that A had benefited more from the condition than had B. This technique makes it possible to compare groups that are made up of subjects who differ on IQ. The technical term for this comparing process is analysis of covariance, a topic beyond the scope of this book. Be aware that this method is available as a means of controlling for variability in an experiment. You may refer to one of the standard books on statistics, such as Kirk (1982), for a description.

**REPLICATION, REPLICATION**

A method of control seldom described as such is replication—the repeating of an experiment to see if the same results are found the second time. Laypersons sometimes assume that once a result has been found by a scientific experiment, the conclusions are fixed permanently. The truth is that an experiment seldom stands by itself, particularly if the results are surprising. In fact, an unusual result remains in a kind of limbo until other experimenters have successfully replicated the experiment. If the same results are obtained by other experimenters, they become part of our scientific knowledge. If the replication is not successful, the supposed facts found in the original experiment are invalid and are forgotten. Many examples of this process have occurred throughout the history of psychology, as well as in the other sciences.

A particularly good example of how dubious phenomena get weeded out is provided by the history of research on transfer of memory by injecting material from a trained animal into an untrained animal. The line of research began with the finding that feeding trained planaria (flatworms) to untrained planaria resulted in a transfer of memory to the untrained worms (J. V. McConnell, 1962). Eventually a similar experiment was tried on rats (Babich, Jacobson, Bubash, & Jacobson, 1965), except that instead of feeding the trained rats to untrained ones, the experimenters injected extracts of their brains. The resultant finding created much interest in the scientific community because of its enormous implications for the mechanisms of memory and for the storage of information in the nervous system in general. A number of
positive replications were published (see W. L. Byrne, 1970, for a review). However, it soon became clear that all was not well with this supposed phenomenon. Not all investigators could replicate the finding. Within a few years of the original research, an article was published by 23 authors from seven laboratories (Byrne et al., 1966) reporting that all of these scientists had failed to replicate the original finding. They stated their conclusions cautiously: "Our consistently negative findings... indicate only that results obtained with one method of evaluating this possibility are not uniformly positive" (p. 658).

In the intervening years, the phenomenon has not been firmly established in spite of continued work. One textbook (Cotman & McGaugh, 1980) summarizes the situation as follows: "In general the findings are extremely conflicting, and as a consequence no firm conclusions can be drawn. Research has not as yet specified either optimal or reliable procedures for producing a transfer effect. Further, it is not at all clear what type of molecule might be responsible for producing the effect... Should such experiments be reproducible, it should be possible to determine the basis of the effects. At the present time, the memory transfer effect must be regarded as not yet convincingly demonstrated" (p. 313).

The history of the research on the transfer of training by injection followed a typical pattern. The first reports elicited a great deal of interest and many attempts at replication. Some of these attempts were successful, and the investigators naturally published their results. Those who did not find the effect were reticent about admitting their failures and, additionally, may have had difficulty getting negative results published. Enough negative results eventually accumulated, however, to overcome biases, and the literature began to reveal a preponderance of negative results. Such a scenario has occurred repeatedly in science, with the result that those research effects that are not repeatable are discarded.

Two types of replication are commonly distinguished: direct and systematic. Direct replication occurs when someone repeats essentially the identical experiment in an attempt to obtain the same results. Systematic replication occurs when Researcher B says, "If A's theory is correct, then the following should happen." Then B performs an experiment different from A's but based on it. If A's results and theory are correct, B should find a certain result.

Direct replication is seldom carried out because finding exactly the same thing as someone else did brings little glory. More specifically, it is difficult to get grants for replications. Journals tend to avoid publishing such research, and professors who spend time replicating other people's work do not get promoted. Direct replication is usually attempted only when systematic replication has failed. Investigators then go back and repeat the original method more exactly in order to pinpoint the source of the difference in results.

Systematic replication is the usual way that experiments are replicated. Researcher B will do an experiment similar to Researcher A's but with
different types of subjects, or with different values of the stimulus, or with different ways of operationally defining the theoretical concepts. All of these approaches are considered systematic replication. As long as results consistent with A’s are found, A’s original experiment is supported by B’s work. You will notice that systematic replication tests external validity by using different subjects, species, or situations. Construct validity is tested when different ways are used to operationalize the theoretical concepts. Statistical validity is tested in all replications, both direct and systematic.

The need for replication is sometimes downplayed in favor of showing that a given result would be unlikely to occur by chance alone. Believers in ESP point out that particular experiments produced results that would have happened only once in billions of experiments by chance alone. The ESP believers’ statistics are usually impeccable, but their understanding of the methods of science is faulty. Innumerable ways exist in which an experiment can fail to be valid, giving results that are due neither to chance nor to the particular hypothesis. Calculating long odds is impressive, but it is only one of a number of considerations in evaluating the experiment. Generally, experiments in ESP fail to replicate. Although believers in ESP may propose reasons for this failure, scientists will pay little attention to ESP until someone devises an experimental situation that gives consistent results in its favor. R. A. Fisher, who largely invented modern statistical methods, said, “Very long odds . . . are much less relevant to the establishment of the facts of nature than would be a demonstration of the reliability of the phenomena” (quoted by Crumbaugh, 1966, p. 527).

**EXPERIMENTAL DESIGN AS PROBLEM SOLVING**

The rest of this book consists largely of examples of good research design. We will not discuss all possible designs, for the simple reason that doing so would be impossible. Rather, we will give a list of designs for you to use as models in designing similar experiments of your own. Experimental designs should be tailor-made for each experimental problem. Sometimes an existing design will fit the problem perfectly. More often, alterations must be made. Therefore, it is better to create the design from the beginning.

Designing an experiment is a matter of solving particular problems of validity by the application of particular methods of control. When every problem has been solved, the experiment is designed. Then is the time to look in books on experimental design to see if your design can be analyzed according to accepted statistical procedures. To look in the books first is to get the cart before the horse and to forget that the essence of experimental design is solving threats to validity in the best way possible.

Two of the general strategies listed at the beginning of this chapter—using the setting as a preparation and instrumentation—are not usually discussed as such in experimental-design books. They are the guiding
principles of design, however. Use the specific strategies we have discussed as tactics in applying these general principles.

THE ELEGANT EXPERIMENT

The goal of every scientist is to design the best possible experiment. How is such a concept put into practice, though? Do you keep testing more and more subjects until the conclusion is inescapable? Do you keep adding variables until every possible source of confounding is taken into account? I find the concept of the elegant experiment helpful in thinking about such questions. In everyday usage, the term elegance implies richness combined with tasteful simplicity. In mathematics, the term emphasizes simplicity. An elegant proof draws a powerful conclusion in the simplest possible way. This idea is what I mean by the elegant experiment: the simplest experiment that will make a clear and convincing test of a hypothesis.

It is possible to include so many variables in a study that not enough measures are made on any one to draw firm conclusions. It is possible to have such a complicated design that you lose sight of the forest for the trees. Consequently, it is important to realize that many trade-offs must be made in the course of designing an experiment and that this process requires hard decisions. Do you spend more time and effort at the outset testing pilot subjects and refining your experimental procedure? Or do you decide to test more subjects in the main experiment to make up for the uncontrolled variability? I cannot tell you what to do in any particular case. Paying careful attention to these questions, though, will result in experiments that are convincing tests of hypotheses.

In selecting the word elegance, I have deliberately chosen a term that has an aesthetic connotation. Designing experiments is an art that requires creativity and that reflects the tastes of the experimenter. Such activity can be both challenging and rewarding.

HOW TO USE THE REST OF THIS BOOK

By now we have an idea of what science is. We have talked about the basic principles of research design, as well as about the principles of validity and the basic means of controlling for threats to validity. We might stop here and tell you to begin designing your own research based on these principles. Obviously we have not done that, and for a good reason. Important as the principles are, probably no one could become a successful researcher by reading a book on the principles of research.

One learns to do research by studying examples of research and, better yet, by doing research. Scientific research is one of those activities that is best learned by working with someone who serves as a guide in a hands-on
situation. In this respect, the way to become a scientist parallels the way an apprentice becomes skilled by working under the direction of an experienced person. The importance of this process can be seen in the many famous scientists who were students of other learned scientists. Firsthand experience in the laboratory of a good scientist has no substitute for learning how to do science. Myriads of attitudes, skills, and techniques are assimilated in such a situation. No book, including this one, can do more than serve as a pale substitute.

My goal in this book is to present those concepts that I spend the most time explaining when I talk with students about research. Although I cannot anticipate all questions, I have tried to answer those that are commonest and most important. Most of you will be taking a course in research methods. Your instructor, along with this book, will be guides as you learn to do research.

* NUTS & BOLTS *

**Choice of Method**

Once you have a question that you wish to investigate, you are faced with many decisions: what kind of subjects, what task, what apparatus, what kinds and values of independent variables? By the nature of this book, we must deal with these questions in a rather general fashion. Nevertheless, certain principles can be stated.

Your review of the literature will reveal standard tasks, apparatus, subjects, and so forth that are generally used in studying a certain problem. For practical and theoretical reasons, it makes sense to follow the standard practice as much as possible and to deviate only when there is good reason to do so. You may feel that a different task, for example, might be more appropriate. What you want to find out will dictate many of these choices. Recall our earlier discussion of the research setting as a preparation. You should make your choices of subjects, apparatus, and so forth with the following question foremost in mind: Which alternative will permit the most sensitive test of the hypothesis? Above all, you must choose a design that will yield data that you can analyze statistically.

**Choice of Subjects**

We will make only general statements about choice of subjects because many considerations are specific to particular studies.

The allegation that the college sophomore and the white rat have been studied too often is true. The reasons become obvious when you start to consider your own alternatives. If the problem lends itself to study with animals, and other experimenters have used the rat, those facts become good reasons for you to do so. Changing strains of rat or even the supplier of the rats can make comparing results between experiments difficult. In addition, using another
species of animal may require you to solve new housekeeping problems. For example, can you keep another species of animal healthy in a laboratory?

If you are studying humans and you wish to use a population other than the college student, what problems will you face with recruitment and payment of volunteers, standardization of technique on the population, and so forth? Such considerations help to explain the extensive reliance on these two classes of subjects. Of course, if you have good reason to use a different population, a little care and effort in designing the study around other subjects often pays big dividends.

**Selection of Subjects**

Ethical and practical considerations enter into the selection process. With the exception of naturalistic observation and certain other types of research, the consent of the subjects must be obtained before they participate. We will discuss the ethics of research in Chapter 14, but meanwhile let us note that subjects in psychological research ordinarily should participate voluntarily.

Ideally, subjects should be a random sample of the population to which you wish to generalize the results of your study. For example, in order to generalize the results of an experiment on college students to the entire adult population of North America, the students ought to be a random sample of that population. Obviously they cannot be, but they should be at least a random sample of college students.

In actuality, most experiments on humans draw subjects from introductory psychology classes. Furthermore, subjects usually volunteer under coercion, for course credit or to meet a departmental requirement. Under these circumstances, you can see that students who sign up for experiments early in the term may be different from those who wait until the last week of class to volunteer. Even going into a class and asking for volunteers will produce a biased sample. For example, women are more likely to volunteer for an experiment when they are in the ovulatory phase of their menstrual cycle and less likely to volunteer when they are menstruating (Doty, 1975). If this variable were important in an experiment, atypical results would be found using women who volunteer spontaneously. The problem can be minimized by calling women randomly from a list of potential volunteers and asking them to come in at a particular time. In this situation they are less apt to participate differentially than if they have volunteered spontaneously.

Your college will likely have standard procedures for recruiting subjects. In fact, all details of the experiment must be approved by the appropriate authorities. You should follow these procedures carefully. For example, if all introductory psychology students must participate in a certain number of hours of experimentation as a course requirement, and if they are supposed to sign up through a central office, then going directly to the classroom to seek volunteers is not fair to other experimenters.

However subjects are to be recruited—in class, by poster, or through newspaper ads—make sure they know the exact building, room, date, and time
of the experiment. You should give them a telephone number to call if they must cancel, and you should have their number in case you must cancel.

Random assignment of subjects to conditions is essential. You must decide in advance your procedure for achieving this. If the experiment is a between-subjects design, you should have a random order made up before the subjects arrive. Then you will follow the assignment as the subjects show up: the first one will receive the condition that is scheduled first, and so forth. Leave no room for subjectivity when you assign subjects to conditions. If subjects are to be tested in pairs, flip a coin to determine which one is assigned to which condition.

How many subjects should you test? There is a rational way of deciding how many subjects to use in an experiment, provided you know how much variability to expect in your data. In ordinary laboratory experiments, however, almost nobody uses this basis for deciding the number of subjects. The reason is simple and practical. Suppose that you wish to achieve a particular degree of precision in your results. Of course, the more subjects you have, the less the means of your data will deviate from their true values. The usual way of representing this error of measurement is called the standard error of the mean. The following equation shows how the standard error of the mean decreases with increases in the number of subjects:

\[ \sigma_x = \frac{\sigma_x}{\sqrt{N}} \]

In this equation, sigma represents the standard deviation of the scores and N is the number of subjects. You can see from the equation that in order to cut the standard error of the mean in half, you must double the square root of N. In order to double the square root of N, you must quadruple N.

How the number of subjects affects the precision of an experiment is illustrated in Figure 6.3, which shows how the standard error of the mean decreases with N. In this figure, we assume that the standard error of the mean is 1 unit when there are 10 subjects. In order to reduce the standard error to the mean to 0.5 unit, we must increase the number of subjects to 40. If we wish to reduce the standard error of the mean again by a factor of 2, to 0.25, we must use 160 subjects!

Thus, you can see that increasing the number of subjects does not decrease the error of measurement in a linear way. Doubling the number of subjects reduces the standard error of the mean by only 30%. (Not all experiments will use the standard error of the mean in data analysis, but the effect is the same for other statistics.) The law of diminishing returns operates. For this reason, most experiments will use about 10 subjects—or, if they have more than one condition, 10 subjects per condition. Observe how many subjects have been used in experiments similar to yours and how much precision was obtained. If you want more precision, ask yourself if using more subjects will make a significant improvement. If not, try to increase the precision of measurement experimentally.
SUMMARY

1. The fundamental meaning of the term control in psychology is that of providing a standard against which to compare the effect of a particular variable.

2. Experiments in which different groups of subjects experience different conditions are known as between-subjects experiments. Those in which each subject experiences every condition are known as within-subjects experiments.

3. The group in a between-subjects experiment that receives the treatment is called the experimental group; the group that does not receive the treatment is called the control group.

4. In within-subjects experiments, the condition that does not contain the experimental manipulation is called the control condition.

5. It is not necessary to have a control group or a control condition in an experiment as long as there is some group or condition that can serve as a comparison for the particular experimental manipulation.

6. A second meaning of the term control is the ability to restrain or guide sources of variability in research. This meaning is captured in the term experimental control.

7. There are three general strategies for achieving control in research: using a laboratory setting, considering the research setting as a preparation, and instrumenting the response.

8. Laboratory research is defined not by the use of a particular kind of room but by the ability to control the important sources of variability in the research setting.
9. The concept of a preparation emphasizes choosing the best possible research situation in which to test a hypothesis.

10. Instrumentation of the response refers to the means of measuring the dependent variables. Careful measurement renders responses objective and may even be thought of as creating responses.

11. Specific control strategies include using subjects as their own controls, randomizing, matching, building nuisance variables into the experiment, and using statistical control.

12. Subjects may be used as their own controls when doing so is logically possible, when serving in all conditions will not destroy their naivete, and when there will not be serious contrast effects between conditions.

13. The allocation of subjects to conditions is random when each subject has an equal chance of being assigned to every condition.

14. Matching may be used when there is an important variable on which subjects differ that is correlated with the dependent variable and where it is feasible to present a pretest to the subjects.

15. Nuisance variables that cannot easily be removed from the experiment may be controlled for by making them independent variables in the experiment.

16. Statistical control may be thought of broadly as synonymous with inferential statistics. More narrowly, statistical control involves equating subjects on paper by means of the analysis of covariance.

17. One of the most important means of control is the replication of an experiment. Direct replication is repeating essentially the same experiment. Systematic replication is doing a different experiment in which certain results should be found if the original experiment was valid.

18. Controlling the sources of invalidity is essentially a matter of solving problems. When the problems are solved, the experiment is designed. The goal of a researcher is to design the most elegant experiment that will answer the questions of interest and will deal with the problems of validity.

19. The choice of method is dictated by factors such as the exact hypothesis to be tested, the methods that are standard in the particular field, and practical considerations.

20. The number of subjects to be used depends on the size of the effect and the anticipated variability of the data. The power of the experiment increases proportionately with the square root of the number of subjects.

Suggestions for Further Reading

Vision in Babies

To illustrate the considerations involved in deciding on the number of subjects to test in an experiment, consider the following example. Professor Strauss studies vision in babies. She wants to find out if they can recognize their mother's face. Her experimental situation involves having a baby look at a pair of faces projected on a screen for 5 minutes. She measures how long they look at the mother's face compared with a control face. Previous studies have shown that if babies recognize their mother's face, they will look longer at the mother than at the control woman. From pilot work, Professor Strauss estimates that they will look 10 seconds longer at the mother's face. The standard deviation of time spent looking is 100 seconds.

Professor Strauss has a budget of $1,000 for this experiment. It costs her $10 to run a single subject. Her apparatus has some problems, however. She believes that a new projector would make the pictures much clearer, increasing the difference in the babies' viewing time for the mother's face to 20 seconds. A new projector would cost $250. Also, her method of recording the babies' preferences has some problems because it is hard to see where the baby is looking. An automatic recording device would reduce the standard deviation of the responses from 100 to 50 seconds but would cost $500.

Professor Strauss scratches her head and wonders what to do. If she uses her old equipment, she can run 100 subjects. If she buys both the projector and the recording device, she can run only 25. Of course, she could buy just one or the other. What should she do?

For this example, the t test is the appropriate statistic. The question then becomes what combination of effect size, standard deviation, and number of subjects will give the largest predicted t.

\[
t = \frac{\bar{X}}{\sigma/\sqrt{N}}
\]

Consider the effects of the various alternatives on this equation.

First, let us use our intuition. Increasing the effect size, decreasing the variability of the response, and increasing the number of subjects should all increase the predicted level of significance of the study.

Now, let us look at the equation and see how this works. Increasing the number of seconds that the babies look at the mother's face compared with the control face will increase the numerator, \( \bar{X} \). So increasing the effect size will increase t. The denominator of the fraction is itself another fraction. The numerator of the bottom fraction is the standard deviation of the number of seconds by which the babies prefer the mother to the control. Decreasing the standard deviation will increase t because it affects the denominator of the main fraction. On the other hand, increasing the number of subjects will increase the value of t because it is the denominator of the lower fraction. Increasing the square root of N will decrease the value of the ratio \( \sigma/\sqrt{N} \). Because this forms the denominator of the main fraction, t is increased.
Determining which alternative is best requires Professor Strauss to work out the value of t for each alternative.

### READING BETWEEN THE LINES

#### 6.1 BRAIN DAMAGE SOMETIMES PRODUCES OBESITY IN RATS

Lesions in a part of the brain known as the ventromedial hypothalamus have been known to produce obesity in rats. Some investigators, however, were unable to find the effect as reliably as others and proposed that the lesion itself did not produce the obesity. Rather, the scar tissue that resulted from the lesion stimulated a nearby area that actually controlled eating. Experiments were done to produce more or less scar tissue and, therefore, more or less irritation to the nearby area of the brain. These experiments showed that the manner in which the lesion was produced did not matter as much as which laboratory did the experiments. One group of investigators consistently found that the lesions produced obesity, while others tended to find no effect. Eventually, a simple difference in methods between the successful and unsuccessful experiments was found. Can you guess what it was?

#### 6.2 COGNITIVE AND AROUSAL FACTORS IN EMOTION

One of the most influential experiments in social psychology was conducted by Stanley Schachter and Jerome Singer (1962), who hypothesized that people experience particular emotions as the result of a cognitive interpretation of a physiological arousal. According to this hypothesis, the same arousal could be experienced, for example, as euphoria or anger depending on the person’s cognitions.

The researchers tested their hypothesis by injecting subjects with a stimulating drug, then exposing them to a confederate who acted either euphoric or angry. The authors reported that the emotions of the subjects tended to match those of the confederate. Later the experiment was attacked in two separate research papers. Because the flaw in the experiment is not obvious, I will tell you that it has to do with replication.

### EXERCISES

#### 6.1 RANDOM ASSIGNMENT

You are designing an experiment in which 10 subjects will experience the experimental condition and another 10 will be given the control condition. A total of 20 subjects have signed up for the experiment. Randomly assign them to the two groups. Describe your steps.

#### 6.2 READING A RESEARCH PAPER

Read the sample paper on pages 341–356 of the text, and answer the following questions.
INTRODUCTION

1. What problem, question, situation, or observation led to this study?
2. What theory is being tested, or what is the theoretical framework within which the work is developed (if it is explicitly mentioned)?
3. What hypotheses or predictions are made, or what will this study contribute to our knowledge of this problem?

METHOD

1. Subjects
   a. Who were the subjects?
   b. How were they selected?
2. Design
   a. What kind of study was this (true experiment, quasi experiment)?
   b. How many conditions were there, and what were they?
   c. Was this a between- or within-subjects design?
   d. How were subjects assigned to conditions (in case of a true experiment)?
   e. Were subjects run blind/double-blind?
   f. Was there a manipulation check (to see if the instructions were effective in setting up the appropriate conditions, in social psychological experiments)?
   g. What were the independent variables?
   h. What were the dependent variables?
   i. Were the subjects debriefed?
   j. Summarize the design.

RESULTS

1. Were any transformations performed on the data?
2. How were the results analysed (what statistical analysis was used)?
3. Was the manipulation check effective (if applicable)?
4. Were the data collapsed over any independent variables, and if so why?
5. Summarize the major findings.

DISCUSSION

1. Were the hypotheses confirmed?
2. How do the findings relate to what is already known about the topic, or what are the implications of these findings?
3. What limitations do these findings have?
4. What future research is suggested by these results?