

7

Experimental Design Independent-Groups Designs

Objectives

After studying this chapter, students should be able to

- Discuss why we do experiments and identify the steps to follow when conducting an experiment
- Compare controlled experiments and field experiments and explain the advantages and disadvantages of each
- Propose an experiment using a completely randomized and a randomized factorial design



A few years ago, one of the authors of your book read an article in a local newspaper with the headline “Don’t Take Engineering, Young Ladies, If You Hope to Marry!” The writer described some data obtained from a major university that indicated that female engineering graduates were less likely to marry than female graduates from other faculties. The reader was left with the impression that women were somehow dooming themselves to singlehood by enrolling in the faculty of engineering. We continue to be amazed to read about the many causal interpretations that are made about data that simply do not permit such interpretations.

We can confidently conclude that there is a cause-and-effect relationship between variables *if and only if* the appropriate study has been conducted. The conclusion that being educated as an engineer causes a decrease in marriageability *could* be made if there was a significant difference between the postgraduate marriageability of two groups of women who were *initially* equivalent in marriageability and were then

randomly assigned to an engineering group and a nonengineering group. And, of course, it would have to be established that during the course of the women's education, variables extraneous to the education experience did not differentially affect the groups. Do you think this was the case? We don't, either.

Why We Do Experiments

The **experiment** is the cornerstone of scientific research. The goal when we conduct experiments is to show that an independent variable (IV) causes a change in the dependent variable (DV). In psychological research, the DV is usually some measure of behavior. Perhaps we would like to know whether technologically enhanced courses as opposed to traditional course delivery techniques (IV) improve student performance in courses (DV). Or perhaps we are interested in comparing psychotherapy with medical therapy (IV) in the treatment of anorexia symptoms (DV). Conducting an experiment is considered to be the best way to provide us with the answers to these kinds of questions. Not all problems lend themselves to experimental study, but those that do are best approached this way. We will begin this chapter by discussing the basics of the experimental approach.

To be a true experiment, as opposed to a quasi-experiment or a nonexperiment, the IV must be *under the control of the researcher*. In other words, the researcher must assign participants to the levels of the IV. Consider the example about psychotherapy versus medical therapy in the treatment of anorexia. If our goal is to compare psychotherapy with medical therapy for the treatment of anorexia, we could study people who had been treated with either therapy in the past. We could determine how well they have progressed with each type of treatment. This approach, however, is not an experimental approach, because *we* did not assign the patients to the type of treatment (IV). Rather, we compared those who *themselves* had chosen either psychotherapy or medical therapy. This type of study is called a *quasi-experimental design*, and such designs will be discussed in Chapter 10. For us to conduct a true experiment to study this problem, we, the researchers, must be able to assign the participants to each condition or group in the experiment. We must be the ones to decide who gets psychotherapy and who receives medical therapy. This is crucial because we, the researchers, can then take steps to ensure that there are no systematic differences between the groups before the experiment begins. As a result, we are able to conclude that if we find a difference between the groups at the end of the experiment, that difference is due to the way the people were treated. Imagine that only the most severely affected individuals got medical treatment, and those with only mild symptoms chose psychotherapy. In this case, the groups were different at the outset of the study; it would be no surprise to find that they are still different at the end of the study.

As we have said, the true experiment is the foundation of scientific research. Although not all research problems can be studied experimentally, when they can, such an approach is preferable because causal statements about the relationship between variables can be made.

CONCEPTUAL EXERCISE 7A

For each of the following hypotheses, decide whether a true or a quasi-experiment is indicated:

1. Young offenders have poorer impulse control than nonoffending youth.
2. Children who view a film promoting helping behavior show more altruistic behaviors than children who view a neutral film.
3. Pigeons on an intermittent schedule of reinforcement exhibit pecking behavior that is more resistant to extinction than pigeons on a continuous schedule.
4. Women rate pornography as less interesting than men do.



Steps in Conducting an Experiment

Step 1. Formulate a Hypothesis

You will recall, from our discussion in Chapter 4, that a hypothesis is a statement about the expected relationships between variables. For example, perhaps we are interested in whether practice in mirror drawing with one hand might transfer to the other hand. In mirror drawing, the participant, while looking in a mirror, attempts to duplicate a figure, number, letter, and so on. Our hypothesis might be as follows:

Positive transfer to the nonpreferred hand will occur with training in mirror drawing with the preferred hand.

As you can see, this hypothesis is a statement about theoretical concepts (i.e., positive transfer and training). The next step is to decide how to measure these concepts. In other words, we have to operationalize, or make measurable, the variables.

Step 2. Select Appropriate IVs and DVs

In the example given above, the IV is amount of practice (practice vs. no practice) with the preferred hand. There are various ways we could measure (or operationalize) positive transfer. One way would be to count the number of errors on three trials before and after practice. Another way would be to measure the time needed to complete a trial. Of course, there are questions to answer. What is an error? How much practice?

Deciding which dependent measure is the most valid and reliable way to measure the behavior of interest is a matter of experience and familiarity with the available research. For more detail on measurement, see Chapter 5.

Once we have a testable hypothesis about the expected relationship between our IV and DV, we need to consider what other variables might be involved and find ways to control them.

Step 3. Limit Alternative Explanations for Variation

Remember, the goal of experimentation is to determine cause and effect. Does manipulation of the IV have a causal effect on the DV? Clearly, there is more than one variable influencing behavior at any one time. Let's look at an illustrative, if somewhat obvious, example. An angler wants to know which type of bait is best for trout. *Type of bait* then will be the IV. The *number of trout* caught in 4 hours will be the DV. Our angler goes to a lake and fishes from 8:00 a.m. until noon on Sunday using flies. The following Sunday, he goes to another lake and fishes with worms from 2:00 p.m. until 6:00 p.m. He catches one trout at the first lake and six at the second lake. If we can demonstrate statistically that more fish were caught at the second lake, we need to ask ourselves, have we demonstrated a *causal relationship* between the IV (type of bait) and the DV (number of trout caught)? Are there alternative explanations for the difference in the number of fish caught that have nothing to do with the IV, type of bait, and therefore are confounded with it? We think so. The second lake might have more trout. The weather on the first Sunday might have been less conducive to fishing. Perhaps fish are more likely to bite in the afternoon than they are in the morning. We are sure you can think of other possible confounds. In an experiment (or any research), we want to control as many of these other variables as possible. Once we have done our best to think about, and limit, alternative explanations for our hoped-for effect, we can go to Step 4.

Step 4. Manipulate the IVs and Measure the DVs

In other words, carry out the experiment. We now have our data. What next?

Step 5. Analyze the Variation in the DVs

In the ideal experiment, all the variation in the DV between groups receiving the IV (i.e., treatment groups) and those not receiving the IV (i.e., control groups) should be caused by the IV. The objective of the exercise is to decrease variation among groups that is not a result of the manipulation of the IV (i.e., error variation). Techniques for reducing error variability have been discussed in greater detail in Chapter 4.

We now must choose the appropriate statistical technique to analyze the variance in the DV. Which procedure we select depends on the kind of data we have and the questions we wish to answer. In Chapter 2, we discussed common analyses used by researchers in psychology today, and in Chapter 13, we will discuss common statistical procedures students might use for their projects. Once we have selected and conducted the appropriate statistical analysis for our data, we can complete the final step.

Step 6. Draw Inferences About the Relationship Between IVs and DVs

We use inferential statistical procedures to make statements about populations based on our sample findings. Conducting a true experiment allows us to make causal

statements about the relationship between the IV and the DV. We can be confident in saying that the manipulated IV caused the changes in the behavior that we measured if we have carried out our experiment carefully by controlling other variables that could provide an alternative explanation of the changes in the DV, leaving our IV as the only likely causal variable.

Where We Do Experiments

Experiments can be conducted in the laboratory (**controlled experiment**) or in a natural setting (**field experiment**). In both controlled and field experiments, the IV is directly manipulated by the researcher. However, in a natural setting, which is where field experiments are conducted, it is more difficult to control all the secondary variables that might affect the results. As a consequence, many researchers prefer to do their work under laboratory conditions if they can. As you can see, there is a downside and an upside to conducting your experiment in both settings.

Controlled Experiments in the Laboratory

There are three broad advantages to controlled experimentation. First, our ability to control the IV is superior under laboratory conditions, improving internal validity. Second, we have superior control over secondary or extraneous sources of variation in the laboratory. For example, we can control noise and temperature. Last, we can more precisely measure our DV under laboratory conditions. This kind of control over the IV, the DV, and secondary variables improves the internal validity of the study—that is, the probability that any changes in the DV are indeed a result of the manipulation of the IV.

Although the experiment is considered by researchers to be the best way to determine cause and effect, there are disadvantages to controlled experimentation. Some phenomena cannot be studied in a laboratory at all. The effect of terrorist attacks on the frequency of travel by Americans, for example, is a research topic that does not easily lend itself to laboratory study. Other research topics present ethical problems. For example, it would be unethical to conduct a laboratory study of the effects of sensory deprivation on human infants. There are also practical disadvantages to laboratory investigation. It can be costly and time-consuming.

But perhaps the most serious disadvantage to controlled experiments is that the outcomes may not be applicable to the real world. For example, Rowland (1999) discusses the validity of studying the human sexual response in the laboratory. Just imagine being hooked up to a number of recording instruments with a camera rolling and someone in a lab coat watching you have sex! Decades of laboratory research has produced a great deal of valuable information, but to what extent do these findings generalize to the bedroom? Behavior that occurs in a laboratory may be idiosyncratic to that environment and may not occur in a natural setting. If we decide to do our experiment in a natural setting, we are conducting a field experiment.

Experiments in the Field

When controlled experimentation is not possible or ethical, a field experiment may be the best choice. A field experiment is conducted in a natural setting (the field) where

the experimenter directly manipulates the IV. Imagine that we are hired to determine if domestic violence intervention training for police officers reduces domestic assault. We might randomly assign a group of police officers to receive special training and a control group of officers to receive the standard training offered by the police department. We could then take various measures, such as the number of domestic assaults in areas served by the two groups or the satisfaction of families served by those officers. Suppose we find that there was a difference. Fewer assaults and more family satisfaction occurred in the areas served by the specially trained officers. Although we would like to think this outcome was caused by our IV (training), we would have to be quite cautious in our inference, in part because in this field experiment, as in all field experiments, we have less control over extraneous, or secondary, sources of variation. The police officers would quickly learn that some of them were in a special training group and some were not. Perhaps the specially trained officers tried harder because they knew they were in the special training group, and it was their extra effort rather than the special training that caused the difference. These “clues” that lead participants to guess about the nature of the study and that may change their behavior are called **demand characteristics**. Researchers must try to anticipate what the demand characteristics associated with their study might be and try to limit their influence on the outcome.

Back to our example. Perhaps there was a factory shutdown in the neighborhoods served by the officers who did not receive special training. When people are out of work, domestic problems probably escalate. It is possible that the greater number of domestic assaults in those neighborhoods might be a result of this factor rather than the training factor. In a laboratory situation, it is often easier to control these kinds of variables by using single- and double-blind procedures. In a single-blind study, participants are not aware of the nature of the treatment group to which they have been assigned; in a double-blind study, neither the participants nor the research staff in contact with them know which group they are in.

The outcome we may find in a controlled laboratory setting may not generalize to the external (i.e., natural) world. When you bring a phenomenon into the laboratory, you may be interfering with how it operates naturally. Findings that do not hold true in a natural setting are not externally valid and may be of little interest. On the other hand, as you saw in our domestic violence example, field experiments, which may be more externally valid, have a major disadvantage in that they may lack internal validity because it is much more difficult to control the IV, the DV, and secondary sources of variation.

Choosing the best setting in which to conduct an experiment requires considerable thought. Some things to think about include pragmatic considerations, such as cost, control over variables, and validity considerations. We have discussed validity in much more detail in Chapter 5.

As we often say, it is a matter of balance.

We have discussed the steps that researchers follow and the things they must think about as they consider how best to answer the questions they have. The rest of this chapter deals with basic experimental designs where different participants serve in each level of an IV—*independent-groups* or *between-participant* designs.

CONCEPTUAL EXERCISE 7B

A researcher is interested in the uninhibited behavior she has often observed at rock concerts. She wonders what kinds of variables influence this behavior (the loudness of the band, the proximity of the audience to the band, etc.). Should she consider a controlled or a field experiment? Why?



How We Do Experiments: Independent-Groups Designs

We cannot stress enough the importance of the assumption of initial equivalence of groups in experimental design. If we cannot assume that our groups were equivalent prior to treatment (i.e., before manipulation of our IV), we have no basis at all for any causal inferences about differences after treatment. So how can we assume that our groups are equivalent at the start?

One common way is to assign participants randomly and independently to conditions—an independent-groups or between-participants design. In an **independent-groups** or **between-participants design**, participants are randomly and independently assigned to each level of the IV. Because the participants were independently assigned to groups or levels of the IV, we can feel confident that the groups were initially equivalent. If we then treat each group differently (i.e., the treatment variable) and find that there is a significant difference in the outcome measure (i.e., the DV), we can confidently infer that the outcome, or differences in the DV, was causally related to our manipulation (i.e., levels of the IV).

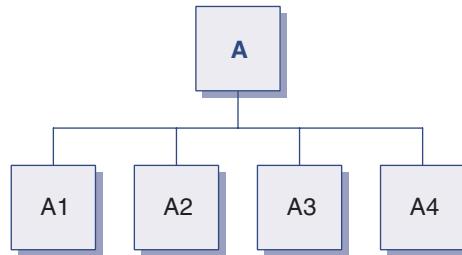
With independent-groups or between-participants designs, each score is independent of every other score, hence the name—independent-groups designs. Because different participants are assigned to the different levels of the IV, their scores are assumed to be independent of each other. The simplest independent-groups design is one where the researcher is interested in one IV with two or more levels—a completely randomized groups design.

Completely Randomized Groups Designs: One IV

In a **completely randomized design**, research participants are randomly assigned to different levels of *one* IV. Figure 7.1 illustrates how we might diagram this type of design.

As you can see, there are four levels of the A IV, and n participants would be randomly assigned to each group.

The simplest completely randomized groups design would be a two-group design where participants are randomly selected and independently assigned to either an experimental group or a control group (i.e., two levels of one IV). Such designs allow

FIGURE 7.1 ● An Independent-Groups Design

us to answer one question: Did the manipulation of the IV affect the DV? Perhaps in more typical language, we might ask, did the treatment of the various groups cause a difference in some response measure, the DV? Let's examine a recently published research article to see how a between-participants design was used in an effort to answer a specific question.



FYI

The American Psychological Association (APA) recommends that the word *participants* be used when referring to human participants in research. Although the word *subjects* has been used, and typically still is used in many books and published research articles,

to refer to various types of research designs, such as between-subjects and within-subjects designs, we have followed the recommendation of the APA in this regard. We use the word *subjects* only when we are referring to animals, not humans.

Randomized Groups Design: One IV With Two Levels

The Research Problem. Childhood sexual assault (CSA) is a traumatic experience that can result in a variety of negative mental health outcomes postassault. Two of the more pressing symptoms include mental health issues such as posttraumatic stress disorder (PTSD) and a higher risk of substance use. Danielson et al. (2012), researchers from the United States, were interested in comparing the effectiveness of a technique called Risk Reduction through Family Therapy (RRFT) with a treatment-as-usual (TAU) control group. Specifically, they wondered if one of these techniques was more effective at reducing substance use and risky sexual behaviors among youth who have experienced CSA.

The Hypotheses. Danielson et al. (2012) hypothesized that both techniques would result in fewer mental health symptoms over time; however, they expected to find that participants in the RRFT group would report less substance use and less risky sexual behavior than those in the TAU group. Their rationale behind this hypothesis is based on how RRFT focuses on these areas and TAU does not.

Selection of Participants and Assignment to Conditions. There were 30 participants (88% female; mean age = 14.8; $SD = 1.5$; range = 13–17) who experienced CSA and were seeking professional help at the outset of the study. The parents of the youth were approached by the researchers to provide consent for their children to participate in the study. The 30 participants were randomly assigned to either the RRFT or TAU group using a computerized blocked randomization method, and a baseline assessment was conducted, along with a urine drug test.

The Independent Variable and Dependent Variables. The IV was the type of treatment received by the two groups of participants, either RRFT or TAU:

- RRFT involves these treatment areas: psychoeducation, coping, family communication, substance abuse, PTSD, healthy dating and sexual decision making, and revictimization risk reduction. Guided by ecological theory, which assumes that there are multiple social influences in the community on an individual's behavior, the RRFT method involves the youth and his or her family in the treatment process.
- TAU is the control group that involved therapeutic treatments that a youth would typically receive when seeking help in alleviating CSA symptoms.

The dependent measures were as follows:

- Mental health symptoms included PTSD symptoms (measured by the UCLA PTSD Index for *DSM-IV*) and depression symptoms (measured by the CDI [Child Depression Inventory]).
- Substance use included drugs and alcohol consumed within the past 90 days (measured by the TLFB [Time Line Follow Back Interview]), which is a common technique used to assess substance use, and confirmed by urine drug screens.
- Risky sexual behavior was measured by the participants' total number of sexual intercourse partners and incidents of STIs (sexually transmitted infections) over the past 3 months.

The Design. Because the participants were randomly assigned to either the RRFT or TAU group and could only be assigned to one condition, this is an independent-groups design.

The Statistical Analysis. The authors took four measurements during the course of the study (baseline, posttreatment, 3-month follow-up, and 6-month follow-up) and analyzed these data using an MRM (mixed-effects regression model) to assess changes over time among the two groups.

The Results. Danielson et al. (2012) reported that the youth in the RRFT group were more significantly impaired at baseline than those in the TAU group, based on MRM results of the UCLA PTSD, CDI, and TLFB, so results should be interpreted cautiously. The RRFT group had significantly greater reductions in PTSD, $\beta(82) = .87, p = .004$, and depressive symptoms, $\beta(81) = .52, p = .036$, compared with the TAU group from baseline to the 6-month follow-up. Significantly fewer days were spent using a substance over time for the RRFT group, $\beta(81) = .30, p < .001$, in contrast to participants in the TAU group. Last, both groups, over time, reported fewer new sexual partners, $\beta(82) = -.01, p = .912$. Although there was not a significant difference between the two groups, $\beta(82) = -.01, p = .912$, the RRFT group expressed having fewer new sexual partners (42% for RRFT vs. 60% for TAU) and fewer STIs (8% for RRFT vs. 27% for TAU).

The Conclusions. Danielson et al. (2012) concluded that, as hypothesized, the RRFT youth had significantly less substance use but not significantly less risky sexual behavior. Interestingly enough, the RRFT group had significantly fewer PTSD and depressive symptoms, although both groups experienced reductions in both of these areas. The authors suggest a replication of their study using a larger sample size to ensure that the two groups are more similar at the outset. Often we are interested in more levels of the independent variable. Here is an example from the literature.

Randomized Groups Design: One IV With More Than Two Levels

For many of us, smartphones have become an essential part of our lives. Can you imagine leaving your house without your phone, losing your phone, or (shudder) dropping your phone in the toilet, a negative experience that at least one of your authors has experienced? For most of us, most of the time, cellular technology has made our lives easier and more productive. But does our love or overuse of these devices have a downside?

On September 1, 2011, the Government of Alberta, Canada, enacted a distracted driving law that strictly prohibits Albertans from using handheld cellular devices while driving (Government of Alberta, 2011). Smartphone use and driving are obviously not a good mix, but what about using your phone when you are just strolling around? Janissa Delzo, in a special report to CNN, described how two young men, completely enthralled by the Pokémon GO games on their smartphones, fell approximately 50 to 90 feet down a cliff in Encinitas, California, when they failed to notice the potential danger of their immediate surroundings (Delzo, 2016).

What about phones in the classroom? Many teachers and university professors (we are two of them) forbid their use during class. We find them distracting to us as we try to deliver our lectures, but we also think they distract our students.

Have you ever used your phone in class? Have you ever wondered if your phone use might affect your grades? Lee, Kim, McDonough, Mendoza, and Kim (2017), from universities and research facilities in the Midwestern and Southern United States, did and decided to conduct a simulated phone distraction experiment to find out.

The Research Problem. In this interesting study, Lee et al. (2017) hoped to determine the influence of three policies governing cell phone use in the classroom (cell phones allowed, cell phones allowed but not used, and cell phones not allowed) on student performance on a 20-item multiple-choice test that immediately followed a 20-minute prerecorded talk. Although the researchers included a fourth no-instruction control, we are only going to discuss the three groups mentioned previously. The talk was a Technology, Entertainment, and Design (TED) lecture on empathy, presented by Dr. Sam Richards. According to the TED website, Dr. Richards is a renowned American sociologist from Penn State University who specializes in race and ethnic relations (Richards, 2010).

Simulated Cell Phone Distraction Experiment. In this independent-groups design, Lee et al. (2017) randomly assigned student participants to each of three groups: (1) cell phones allowed (CPA); (2) cell phones allowed but not used (CPANU); and (3) cell phones not allowed (CPNA).

The Groups. The students who were randomly assigned to the CPA group were told by the researchers that they could use their phones, but only for educational purposes. The students in the CPANU group were told to put their phones on silent mode and not use them during the lecture period, and the students in the CPNA group surrendered their phones to the researchers before the lecture period began. At the end of the recorded lecture, all participants completed the 20-item multiple-choice test.

The Distraction. At 3, 7, 11, and 15 minutes into the lecture period, the researchers sent four consecutive text messages to the students in the CPA and CPANU groups.

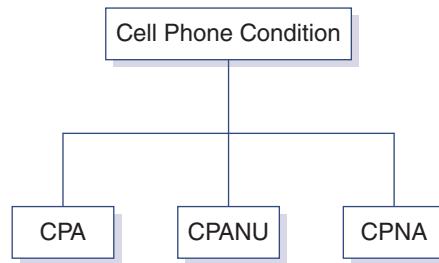
Three minutes after the lecture started, these students got this text: “Hey! Are you there?”

Four minutes later, they got a second text: “We are waiting at the McDonald’s.”

Four minutes later, they got a third text: “Are you coming?”

And four minutes after that, they got the final text: “Sorry! I got the wrong number.” (Lee et al., 2017, p. 362). Lee et al. (2017) were careful to make sure that the text messages resembled the type of text messages typically received from friends of the students and that they were in no way related to the actual material covered in the lecture.

One minute after the recorded lecture ended, participants in all three groups completed a 20-item multiple-choice test based on the material covered in the empathy lecture. Let’s take a look at Figure 7.2 to see exactly what went on in Lee et al.’s (2017) simulated phone distraction experiment.

FIGURE 7.2 ● Independent-Groups Design: One IV > 2 levels

Source: Lee et al. (2017).

The Hypotheses. Lee et al. (2017) had two hypotheses: (1) They hypothesized that students who were permitted to keep their phones (CPA and CPANU) would be distracted by the texts they got on their phones and would do worse on the test than those who had to surrender their phones prior to the lecture period (CPNA) and (2) that students who were not allowed to use their phones during the lecture (CPANU) would do worse on the test than those who were allowed to use their phones (CPA).

Selection of Participants and Assignment to Conditions. Lee et al. (2017) randomly assigned 120 volunteers from introductory psychology classes, who received credit toward their psychology course grade, at a college in Southeastern Arkansas to each of the phone policy groups.

The Independent Variable and the Dependent Variable. The IV was the type of phone use that the students were permitted during the lecture.

Lee et al. (2017) included four DVs in their study, but we are only going to discuss the number of correct answers on the 20-item multiple-choice test that was administered by the researchers immediately following the lecture.

The Design. This study is an independent-groups design with three levels of one IV. See Figure 7.2.

The Statistical Analysis. Because participants were randomly assigned to each of three levels of one IV, the researchers analyzed the DV (participant test scores on the 20-item multiple-choice test) using a one-way ANOVA.

The Results. The descriptive statistics of Lee et al.'s (2017) study are presented in Table 7.1. The scores are the percentage of correct answers on the test.

TABLE 7.1 • The Descriptive Statistics: Means (Standard Deviations)

Cell Phone Condition		
CPA	CPANU	CPNA
53.12 (15.14)	52.88 (13.91)	61.62 (14.87)

Source: Adapted from Lee et al. (2017).

For the inferential analyses, Lee et al. (2017) used a one-way ANOVA. The outcome of an ANOVA, if significant, tells us only that at least two means are different. There may be other pairs that differ. We need to do further post hoc analyses, and Lee et al. (2017) chose Tukey's honestly significant difference (HSD) tests to find out.

The one-way ANOVA revealed that there was a significant difference among the phone conditions, $F(3, 156) = 3.14, p = .027$.*

The outcome of the Tukey test comparisons at an $\alpha = .05$, were as follows:

Students who had to surrender their phones to the researchers at the onset of the lecture (CPNA) significantly outperformed students who were allowed to keep their phones (CPA), and they also outperformed students who were allowed to keep their phones but were not permitted to use them (CPANU).

The Conclusions. The researchers speculated that students who were instructed to surrender their phones prior to the lecture were able to focus more on the task at hand and therefore perform better on the multiple-choice test than the two groups of students who were allowed to keep their phones. These findings support the researchers' first hypothesis but not their second, because the researchers also expected that students' having to surrender their phones or being allowed to keep them but not use them (CPNA and CPANU) would be enough of a distraction to cause those students to do worse on the multiple-choice test than the students who kept their phones and could use them (CPA). This was not the case, however.

If you look at Table 7.1, you can see that students who had their phones but were not allowed to use them (CPANU), did no better on the multiple-choice test than those who were allowed to use them (CPA).

The findings of Lee et al. (2017) should make us all pause and think about how our phones might be distracting us as we go about our daily lives.

Randomized Factorial Groups Designs: More Than One IV

How were your grades in high school? Do you think they accurately reflected your level of achievement? Perhaps your classmates, teachers, or textbooks influenced your grades. Human behavior is complex and influenced by all sorts of variables.

*Note: You may have noticed that the *df* indicate that four groups were involved, as was indeed the case. The authors did not report a mean or standard deviation for the control group.

Experiments designed to assess the effects of more than one IV on performance are probably more like the real world and more likely to be externally valid (i.e., generalize to real-world settings and situations).

In a **randomized factorial design**, participants are randomly assigned to each level of *more than one* IV (or factor). These designs allow us to assess the effects of more than one IV and to assess the interaction between IVs. A *simple randomized groups design* provides the answer to *one* question: Did the IV affect the DV? *Factorial designs* allow us to find answers to *several* questions: What effect, if any, did each IV have on the DV, and how did the combination of levels of the IVs affect the DV? The statistical analysis of a factorial design, then, allows us to assess the effects of each IV on the DV (called the **main effects**), and it will indicate the interactions among those IVs (called **interaction effects**). Interaction effects are very important; indeed, if significant interactions among IVs are present, the interpretation of the effects of those IVs becomes more complicated. Figure 7.3 illustrates how we might diagram a randomized factorial design.

As you can see, this design has two IVs: (1) A has four levels, and (2) B has two levels. Participants would be independently assigned to each of the eight groups.

Before we look at some real randomized groups experiments, let's examine a hypothetical experiment. Many people, including us, find that reading under incandescent light is easier on the eyes than reading under fluorescent light. But is text easier to read (i.e., more legible) under incandescent light? Word processors offer many different fonts. We find certain fonts to be easier to read than others. Perhaps some fonts are easier to read under incandescent light and others under fluorescent light. We could design an experiment to determine if our anecdotal experience with light and fonts is supported empirically. We might randomly assign readers to incandescent light conditions and others to fluorescent light conditions, and we might have them read text typed in three different fonts under each type of light condition. With two levels of our first IV (light) and three levels of our second IV (font), we would have six different conditions in our experiment. If we randomly assigned different readers to each condition, we would have a 2×3 randomized factorial design (i.e., two levels of one IV and three levels of a second IV) that might look like the following (see Table 7.2).

FIGURE 7.3 ● Randomized Factorial Design

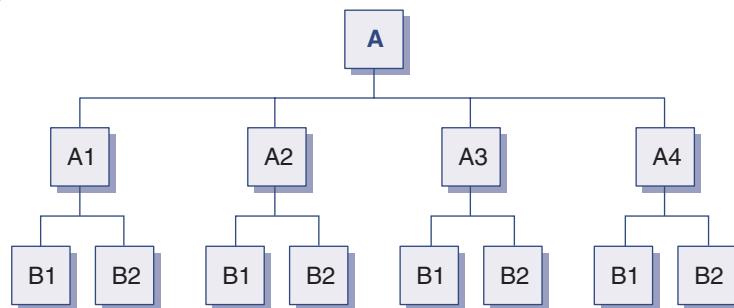


TABLE 7.2 • Hypothetical Randomized Factorial Design Comparing Different Kinds of Light and Font

Font (Second IV)	Light (First IV)	
	Incandescent	Fluorescent
Times	IT	FT
Courier	IC	FC
Geneva	IG	FG

As you can see, readers in the IT condition would read text in the Times font under incandescent light, and readers in the FG condition would read the same text in the Geneva font under fluorescent light. Following the reading part of our study, suppose we had asked our participants to rate the text for readability. In a randomized factorial design, we would have three questions. In our hypothetical experiment, the first question to ask is “Did the first IV (i.e., light) affect the DV (readability)?” In other words, forgetting about the different fonts used for the moment, did light condition affect readability overall? We could examine the mean readability ratings calculated over all font conditions. We would simply determine the mean readability rating for *all* the participants who read the text under incandescent light, regardless of the font used, and compare those ratings with the mean readability ratings of *all* the participants who read the text under fluorescent light. Let’s plot these ratings in a graph. Figure 7.4 shows the main effect of light for our hypothetical experiment. Higher ratings indicate that the text was more readable.

It seems that text is more readable under incandescent than under fluorescent light. In a real experiment, we could analyze these data with a two-way ANOVA to determine if the difference was statistically significant.

Our second question is “Does our second IV (i.e., font) make a difference, overall?” Now, we will look at the readability scores for each level of font calculated over both light conditions. In other words, we will find the mean score for *all* the readers who read the Times font, *all* those who read the Courier font, and *all* those who read the Geneva font, regardless of light condition. Figure 7.5 is a graph of the main effect of font.

Our graph indicates that the Times font is easiest to read, and the Courier font is the hardest to read. Our two-way ANOVA would tell us if the groups differ significantly.

In a randomized factorial design, the main effects must be interpreted in light of any interaction effects. When we examine a main effect, we are looking at means for each level of that IV calculated over all the levels of the other IVs, so we cannot see any differences that might exist at the different levels of the other IVs. The following is our third question: “Is the DV affected by the combination of levels of each IV?”

Let’s now graph the mean readability scores for *each* level of each IV to see if there is an interaction effect (see Figure 7.6).

FIGURE 7.4 ● Main Effect of Light

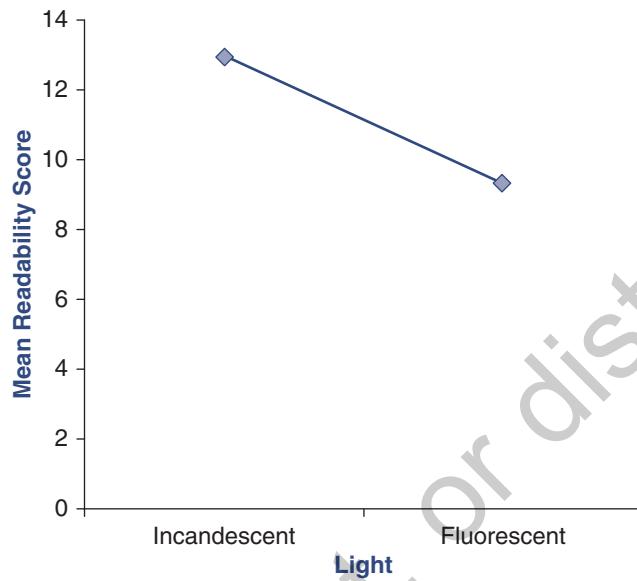


FIGURE 7.5 ● Main Effect of Font

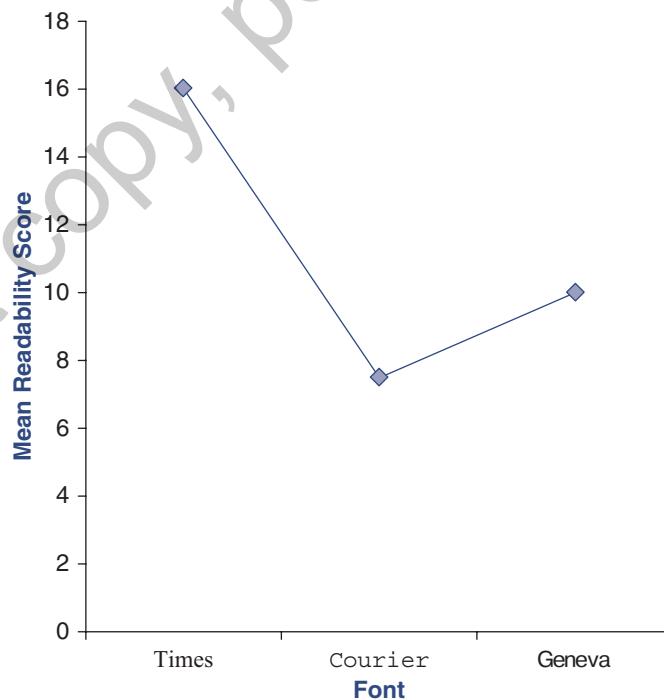
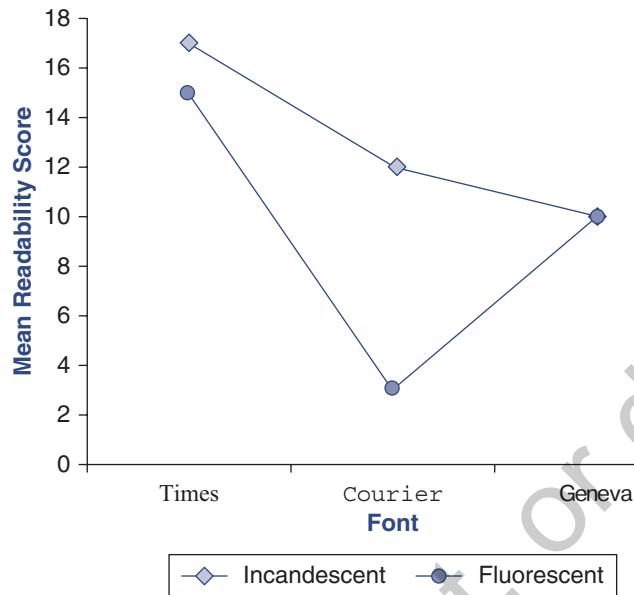


FIGURE 7.6 ● Interaction of Font With Light

You can see that things are not quite as simple as they seemed when we looked at the graphs of the two main effects. If the two IVs did not interact, we would see two more or less parallel lines. These lines are not parallel, so it seems that we have an interaction going on here. Of course, the statistical analysis would tell us if the interaction is statistically significant, but let's examine our graph. Readability of the Times and Geneva fonts seems to be little affected by light conditions. Readability of text in these two fonts is similar, but this is not the case for the Courier font. It seems to be much harder to read the Courier font under fluorescent light than under incandescent light. The interaction effect tells us that the DV (readability) is affected differently under the different combinations of the IVs. This is the great advantage of a factorial design. We can determine how different combinations of levels of two (or more) IVs affect the DV. We could not determine this effect with a simple independent-groups design.

Are you ready for a real experiment?

Have you noticed the frightening messages and images that have appeared on cigarette packages in many countries over the past few years? We sure have. The creators of these warnings were hoping to decrease smoking, but do such negative health messages and images work? Would more positive messages be a better way to change behavior? That is what Hollands and Marteau (2016) of the University of Cambridge decided to investigate, but their focus was on food messages.

They wondered, for example, whether people would spend less money on potato chips if they were confronted with an ugly image of heart disease on the bag.

Perhaps people would buy more apples if a decal of a happy jogger was displayed on each apple?

The Research Problem. Hollands and Marteau (2016) were interested in determining the precise conditions that lead people to make healthy rather than unhealthy food choices. They conducted an online conditioning experiment to find out.

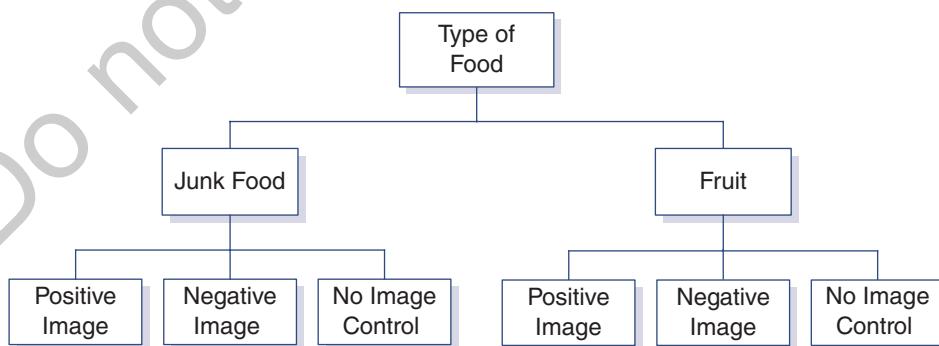
The Online Conditioning Experiment. Hollands and Marteau (2016) delivered their conditioning experiment online by presenting a slideshow to their participants.

In this 2 (junk food or fruit) \times 3 (positive, negative, or no health image) factorial design, the researchers randomly assigned participants to each of six groups. Let's create a diagram so that we can see exactly what went on (see Figure 7.7).

Hollands and Marteau (2016) used five junk food images (chocolate, cookies, cake, potato chips, and a combination of those items) and five fruit images (an apple, an orange, a banana, grapes, and a combination of fruit) paired with five positive health images (images of happy-looking people engaging in healthy activities, such as jogging or cycling), five negative health images (one image of an obese woman, one image of an obese man, two pictures of arterial disease, and one of open-heart surgery), or a blank screen in their online slideshow. Each of the five junk food or fruit items was paired with each of the five positive, negative, or blank screens, and the image pairs were shown 20 times in random order during the online slideshow.

During each trial of the conditioning procedure, participants were shown either a junk food or a fruit image at random on their computer screens for 1,000 ms, followed by the health image (positive, negative, or blank screen) for another 1,000 ms, and finally a blank screen for another 500 ms. Each of the 100 trials lasted 2,500 ms. If you are curious about the total amount of time it took to view the online slideshow, Tanya, our zealous student, who worked with us on this edition, did the math. She reported that at 2,500 ms over 100 trials, participants repeatedly viewed images on their screens for approximately 4 minutes. We are quite sure her math is stellar.

FIGURE 7.7 ● Hollands and Marteau's 2 \times 3 Design



After participants finished viewing the slideshow, they completed a post-conditioning behavioral choice task. They were shown four junk food images (a chocolate bar, cake, chips, and cookies) and four fruit images (an apple, a banana, an orange, and grapes) and were simply asked to make two choices as to which type of food item they most felt like eating. If the participants did not find any of the food items particularly appetizing, choosing none was also an option.

The Hypotheses. Contrary to what the creators of the horrific warnings on cigarette packages seemed to think, Hollands and Marteau (2016) hypothesized that images associated with good health or potential positive health consequences would have a stronger effect on food preference than images of ill health or potential negative health consequences. The researchers also hypothesized that participants who viewed fruit paired with positive health images in the conditioning trials would be more inclined to pick fruit, as opposed to junk food, in the post-conditioning behavioral choice task.

Selection of Participants and Assignment to Conditions. A national research agency recruited 711 people living in England by offering them a £4 (or \$6.00 US) voucher to spend at a UK food store as compensation for participating in this online experiment. Participants were randomly assigned to each of the six online conditioning groups.

The Independent Variables and the Dependent Variable. One of the IVs in this experiment was the health-conscious food item (fruit) versus the less health-conscious item (junk food) that the participants viewed. The negative image, positive image, or no health image control was the second IV in this study.

Hollands and Marteau (2016) included three DVs in their study, but we are going to discuss only the post-conditioning behavioral choice task. In case you forgot, after the slideshow, participants were shown four junk food and four fruit items and were simply asked to make two choices as to which type of food items (junk food, fruit, or no food item at all) they felt most like eating.

The Design. This was a 2×3 randomized factorial design. Participants were randomly assigned to one of the six conditions (two types of food items \times three types of images).

The Statistical Analysis. The appropriate statistical analysis for this design is a factorial ANOVA, and that is what the researchers used. They also used Bonferroni tests in their post hoc analyses to determine specific mean differences.

The Results. The researchers scored participants' preference for fruit as +1, their preference for junk food as -1, and no preference for either junk food or fruit as 0. Participant scores for each of the two behavioral choice tasks were summed on a 5-point scale of preference, with a positive score indicating a preference for fruit and a negative score indicating a preference for junk food.

The descriptive statistics from Hollands and Marteau's (2016) experiment are presented in Table 7.3.

TABLE 7.3 ● The Descriptive Statistics: Means (Standard Deviations)

Type of Health Image	Type of Food					
	Junk Food			Fruit		
	Positive	Negative	No Image Control	Positive	Negative	No Image Control
	-.03 (1.61)	.63 (1.49)	.01 (1.55)	-.04 (1.54)	.44 (1.44)	-.08 (1.62)

Source: Adapted from Hollands and Marteau (2016).

A 2×3 factorial ANOVA was used to assess the participants' preference for either health-conscious (fruit) or less health-conscious (junk food) items in the post-conditioning behavioral choice task. Hollands and Marteau (2016) reported no significant main effect of food type (junk food or fruit) and no interaction between food type and health image. Simply speaking, repeatedly viewing images of junk food or fruit did not influence participants' decisions to pick that food item in the behavioral choice task, nor did any one particular combination or "pairing" of junk food or fruit images with a positive, a negative, or no health image affect participants' food preference for either a health-conscious (fruit) or less health-conscious (junk food) item in the post-conditioning behavioral choice task. The researchers did, however, find a significant main effect of type of health image, $F(2, 705) = 10.20, p < .001$.

The post hoc Bonferroni tests revealed that participants in the two groups who were shown negative images were more likely to choose fruit (a healthy food item) than those in the groups with no image control and those in the groups with positive images.

The creators of those frightening images on cigarette packages may be on the right track!

Hollands and Marteau (2016) found statistically significant evidence that negative health images have a stronger effect on healthy food choice than positive or no health images do, but, as you can see in Table 7.3, it did not matter if negative health images were paired with junk food or fruit; both of these pairings led to an increase in fruit selection (a healthy food choice) over junk food (an unhealthy food choice) in the behavioral choice task. It seems that exposure to negative health images in general leads people to make healthy food choices, whether those images are placed on a bag of potato chips or not. Simple exposure to negative health images seems to be key!

The Discussion. Hollands and Marteau (2016) investigated factors that may lead people to make healthy as opposed to unhealthy food choices.

They thought that participants would be more inclined to choose fruit (a health-conscious food item) when it had been associated with images of good health or potential positive health consequences, but this hypothesis was not supported.

What they did find was that participants were significantly more likely to choose a healthy food item (fruit) after viewing negative health images. It did not matter if those negative health images were paired with junk food or fruit; both led to an increase in fruit selection in the behavioral choice task.

Although Hollands and Marteau (2016) did acknowledge the importance of laboratory-based experimental methods for the purpose of health intervention development, they also cautioned that one of the limitations of using such methods is that they may not provide evidence of the efficacy of these interventions outside of the laboratory setting (i.e., they may lack external validity).

The researchers make a good point. Perhaps people in a grocery store behave differently than people looking at pictures of food on their computer screens.

Hollands and Marteau (2016) did not find an interaction between their two IVs, but had they found one, the main effects would have been interpreted differently. The factorial design allows us to investigate more complicated and often more interesting relationships among variables.

Hollands and Marteau's (2016) 2×3 factorial design involved two variables with two and three levels, respectively. The complexity of a factorial design increases with the number of levels of each variable and with the number of IVs. With only two variables, the design yields two main effects and one interaction effect. With three IVs, the design yields three main effects, three two-way interaction effects, and one three-way interaction effect. As you can imagine, the more variables, the more complicated the interactions, and the more difficult it is to interpret the findings. But given the complexity of human behavior, designs involving multiple variables are the rule in social science research today.

In addition to increasing complexity, the number of participants also dramatically increases with the addition of each new variable. Hollands and Marteau (2016) randomly assigned participants to each of the six online conditioning groups: 711 participants in total. What if they were to add a third IV? Perhaps they also wanted to see if the number or repetitions of the food-image pairings made a difference (low versus high, for example). If they did, they would have to add six more groups to the experiment and recruit a lot more participants.

Independent-Groups Designs: One IV and One Participant Variable

Following people around and/or observing their lives without their knowledge is called *stalking* and is a crime in many countries. Following people around and/or observing their lives on social media without their knowledge is called *creeping* and, as far as we know, is not a crime.

Creeping somebody's Facebook profile is like peeing in a swimming pool: We've all done it, but we don't like to admit it. Muise, Christofides, and Desmarais (2014), from universities in Ontario, Canada, looked at a form of creeping on an imitation Facebook site (IFS) in their study of gender differences in jealousy.

As you know, a true experiment involves true IVs. Muise et al. (2014) included a true IV in their study, but they also included student gender, a participant variable.

The Research Problem. There is evidence that men and women experience, and react to, jealousy in different ways. For example, researchers have found that men experience greater feelings of jealousy as a result of *sexual infidelity* and women experience greater feelings of jealousy as a result of *emotional infidelity* (Buss, Larsen, Westen, & Semmelroth, 1992, as cited in Muise et al., 2014). In addition, once evidence of infidelity is detected, jealous women are more likely than jealous men to increase their surveillance of their offending partners by spying on them, checking up on them, or looking through their things for evidence of betrayal (Pfeifer & Wong, 1989, as cited in Muise et al., 2014).

Jealousy-Provoking Facebook Experiment. In the first phase of Muise et al.'s (2014) study, a research assistant directed heterosexual male and female university students to an imitation Facebook profile (IFP) and told each to imagine that the IFP belonged to him or her and that another profile page linked to his or her profile belonged to his or her partner. Sounds a bit confusing, right? Let's say that you are a student named Jane, and your boyfriend is John. The research assistant would ask you to pretend that the profile you see is yours and that a linked profile is John's.

After the students were acquainted with their own IFPs, the research assistant then directed each student to visit his or her hypothetical partner's IFP.

In this phase of the study, Muise et al. (2014) randomly assigned students to view one of three types of images that depicted their hypothetical partner with an attractive person of the opposite sex: (1) an unknown person (UP), (2) a mutual friend (MF), or (3) the hypothetical partner's cousin (C). Students' hypothetical partners were pictured in a bar setting with the UP, MF, or C (A. Muise, personal communication, October 24, 2017).

The researchers thought that seeing an image of the hypothetical partner with an unknown person (UP) of the opposite sex would be more threatening to students and provoke more feelings of jealousy than seeing an image of the hypothetical partner with a mutual friend (MF) or cousin (C) of the opposite sex. The specific images were chosen by the researchers after pilot testing indicated that the images might lead to different levels of jealousy (A. Muise, personal communication, October 24, 2017).

After participants were exposed to one of the three images, the research assistant advised the students that they were free to explore the IFS, which came complete with photo albums, status updates, and not only their own and their hypothetical partner's profiles but also the IFP of the UP, MF, or C depicted in the image that the student had just viewed.

Two DVs were measured. The researchers measured the total amount of time, in seconds, that each student spent exploring, investigating, or creeping the IFS before logging out. After students logged out of the IFS, they completed a 9-item online jealousy scale.

Selection of Participants and Assignment to Conditions. Eighty-three male and 77 female undergraduate students, recruited from a Canadian university psychology participant

pool, participated in this study. Students were eligible to participate in the study if they were Facebook users, were heterosexual, and had personal experience (either at the time of the study or in the past) being a friend with their partner on Facebook. The mean age of the students was 19.16 years, and the participants were randomly assigned to one of three conditions.

The Independent Variable and Dependent Variables. In this study, the true IV was the type of hypothetical partner image used by the experimenters (UP, MF, or C).

The other variable (a participant variable) was gender. The researchers randomly assigned male and female students to one of three types of hypothetical partner images.

There were two DVs in this study. The first was the total amount of time, in seconds, that students spent exploring, investigating, or, in other words, creeping the IFS before logging themselves out.

The second DV was the score on the 9-item, 7-point jealousy scale that students completed after logging out of the IFS. Students were required to respond to items such as “I would suspect that my partner is secretly seeing someone else” on a scale from 1 (strongly disagree) to 7 (strongly agree; Muise et al., 2014, p. 39). Higher scores indicated more jealousy.

The Design. The experimental design was a 2 (gender) \times 3 (type of image) between-participants design. This is an independent-groups design with one true IV with three levels and a second participant variable with two levels.

The Hypotheses. Remember that all the students viewed images of their hypothetical partner with an attractive person of the opposite sex.

Muise et al. (2014) had two hypotheses about the jealousy scores:

- Students who saw images of their hypothetical partner with an unknown person (UP) would be more jealous than students who saw images of their hypothetical partner with a mutual friend (MF) or with a cousin (C).
- Students who saw images of their hypothetical partner with a mutual friend (MF) would be more jealous than students who saw images of their hypothetical partner with a cousin (C).

Muise et al. (2014) had one hypothesis about the *total amount of time* that students would spend creeping the IFS. The researchers expected gender and type of image to interact. Specifically, they expected that female students who saw images of their hypothetical partner with an unknown person (UP) would spend the most time creeping the IFS, and that female students who saw images of their hypothetical partner with a cousin (C) would spend the least time creeping, with the mutual friend (MF) group in between. The researchers did not expect to see any differences in creeping behavior for the male students.

The Statistical Analysis. Muise et al. (2014) conducted a multivariate analysis of covariance (MANCOVA) in order to determine gender differences (male vs. female) and the effect of type of hypothetical partner image (UP, MF, and C) on the jealousy scale and the total amount of time that students spent creeping the IFS. The researchers also used Tukey (HSD) tests in their post hoc analyses to determine specific mean differences.

The Results. Jealousy: The descriptive statistics for the jealousy data are presented in Table 7.4.

The MANCOVA revealed a significant main effect of gender on the jealousy scores, $F(3, 150) = 16.16, p < .001$. Women had significantly higher jealousy scores than men after logging out of the IFS.

The researchers also found a significant main effect of type of hypothetical partner image on the jealousy scores, $F(3, 150) = 8.87, p < .001$. The Tukey test comparisons revealed that students who viewed images of their hypothetical partners with an unknown person (UP) and those who viewed images of their hypothetical partner with a mutual friend (MF) had higher scores on the jealousy scale than those who viewed images of their hypothetical partner with a cousin (C). However, the mean jealousy scores of students who viewed images of their hypothetical partners with an unknown person (UP) and those who viewed images of their hypothetical partners with a mutual friend (MF) did not significantly differ.

As you recall, the researchers had two hypotheses about the jealousy scores. The first was that students who saw images of their hypothetical partner with an unknown person (UP) would be more jealous than students who saw images of their hypothetical partner with a mutual friend (MF) or with a cousin (C). This hypothesis was partially supported: Although the UP and MF groups did not differ, both groups had significantly higher scores on the jealousy scale than the C groups.

The second hypothesis was that students who saw images of their hypothetical partner with a mutual friend (MF) would be more jealous than students who saw images of their hypothetical partner with a cousin (C). This hypothesis was supported: Students in the MF groups reported significantly more jealousy than those in the C groups.

TABLE 7.4 • Descriptive Statistics for Jealousy in Muise et al.'s (2014) Study: Means (Standard Deviations)

Gender					
Male			Female		
Type of Hypothetical Partner Image			Type of Hypothetical Partner Image		
UP	MF	C	UP	MF	C
2.95 (0.18)	3.72 (0.17)	2.84 (0.18)	4.10 (0.19)	3.91 (0.18)	3.37 (0.18)

Source: Adapted from Muise et al. (2014).

As you know, main effects must always be interpreted in light of any significant interactions. Muise et al. (2014) found an interaction between gender and type of hypothetical partner image on the jealousy scale, $F(3, 150) = 3.65, p = .03$.

Figure 7.8a illustrates this interaction. Let’s take a look.

As you can see, female students in Muise et al.’s study (2014) felt more jealous after seeing images of their hypothetical partner with an unknown person (UP) and with a mutual friend (MF) than with a cousin (C). Male students, on the other hand, felt more jealous after seeing their hypothetical partner with a mutual friend (MF) than an unknown person (UP) or cousin (C). Let’s now see what the researchers learned about creeping time.

Creeping Time: The descriptive statistics for the time spent creeping data are presented in Table 7.5.

FIGURE 7.8A • Gender × Type of Hypothetical Partner Image on Jealousy Score

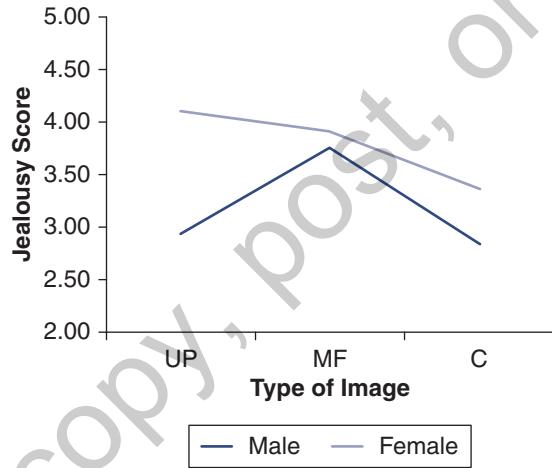


TABLE 7.5 • Descriptive Statistics for Time, in Seconds, Spent Creeping the IFS in Muise et al.’s (2014) Study: Means (Standard Deviations)

Gender					
Male			Female		
Type of Hypothetical Partner Image			Type of Hypothetical Partner Image		
UP	MF	C	UP	MF	C
139.75 (18.43)	79.14 (18.11)	139.78 (18.76)	195.04 (19.90)	137.65 (19.12)	104.48 (18.76)

Source: Adapted from Muise et al. (2014).

The researchers had one hypothesis about the time spent creeping data: The researchers expected gender and type of image to interact. Specifically, they expected that female students who saw images of their hypothetical partner with an unknown person (UP) would spend the most time creeping the IFS, and that female students who saw images of their hypothetical partner with a cousin (C) would spend the least time creeping, with the mutual friend (MF) group in between. The researchers did not expect to see any differences in creeping behavior for the male students.

The MANCOVA revealed a significant main effect of type of hypothetical partner image on time spent creeping the IFS in Muise et al.'s (2014) study, $F(3, 150) = 5.43, p = .02$.

The Tukey test comparisons revealed that students who viewed images of their hypothetical partner with an unknown person (UP; $M = 168.40, SD = 13.56$) spent more time on the IFS site than those who viewed images of their hypothetical partner with a cousin (C; $M = 120.13, SD = 12.27$) or a mutual friend (MF; $M = 108.40, SD = 13.17$).

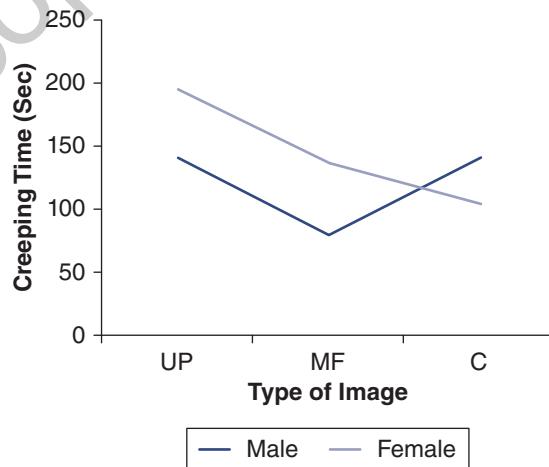
As expected, gender and type of image did interact, $F(3, 150) = 4.18, p = .02$.

Let's take a look at this second interaction in Figure 7.8b.

As you can see, female students spent *more* time on the IFS after viewing images of their hypothetical partner with an unknown person (UP) than with a mutual friend (MF) or cousin (C). This is what the researchers expected.

The researchers did *not* expect to see any differences in creeping behavior for the male students, but, in fact, there were differences. Male students who viewed images of their hypothetical partner with a mutual friend (MF) spent significantly *less* time on the IFS than male students who saw images of their hypothetical partner with an unknown person (UP) or a cousin (C).

FIGURE 7.8B ● Gender × Type of Hypothetical Partner Image on Creeping Time



The Discussion. Muise et al. (2014) investigated gender differences in jealousy and social media creeping in an imitation Facebook experiment.

As the researchers expected, women felt *more* jealous after viewing images of their hypothetical partner with an unknown person (UP) or a mutual friend (MF) than they did after seeing images of their hypothetical partner with his cousin (C), and they spent *more* time on the IFS after viewing the unknown person (UP) and mutual friend (MF) images.

The findings for the men were somewhat puzzling. The men in the MF groups reported feeling significantly *more* jealous than the men in the UP and C groups, but their creeping time did not reflect this: They actually spent *less* time creeping the IFS than those in the other two groups.

Can opposite-sex friendships truly be platonic? We suspect that many people doubt that they can. What do you think?

Between-participants designs, where participants are randomly and independently assigned to conditions, are excellent experimental designs, particularly if there is any concern that one treatment condition might contaminate another. However, the fact that participants have been randomly assigned to conditions does not guarantee the initial equivalence of groups. Nevertheless, the random assignment of participants to conditions is the most common technique researchers use to deal with initial differences between participants. Another approach to ensuring the equivalence of groups will be addressed in Chapter 8.

Chapter Summary

Experiments are the first choice of most behavioral researchers because they allow us to infer a *causal relationship* between a manipulated *independent variable (IV)* and some measure of behavior, the *dependent variable (DV)*. We conduct experiments to evaluate theories, satisfy our curiosity, and demonstrate behavioral phenomena and the factors influencing them.

When we conduct experiments, we begin with a *hypothesis*. We select appropriate IVs and DVs to test our hypothesis. We make every attempt to control alternative sources of variation, and then we carry out our experiment. After analyzing the data, we are in a position to draw *inferences* about the relationship between our manipulated IV and the observed behavioral change.

Experiments can be conducted in a laboratory; these are called *controlled experiments*. When we conduct experiments in a natural setting, they are called *field experiments*. Controlled experiments tend to have greater *internal validity* than field experiments, but field experiments may have greater *external validity*. Some research problems are better examined in a natural setting, whereas others are better examined in the controlled conditions of the laboratory.

Independent-groups or *between-participants designs* are used to compare different groups of participants, all of whom have been independently assigned to treatment groups. The simplest of these is the *completely randomized design* with one IV with two levels. When participants have been independently assigned to all combinations of

more than one IV, we have a *randomized factorial design*. Factorial designs allow the simultaneous assessment of more than one IV and the interactions between IVs. The effects of each IV on the DV are called *main effects*, and the effects of combinations of levels of IVs on the DV are the *interaction effects*.

CHAPTER RESOURCES

Answers to Conceptual Exercises

Conceptual Exercise 7A

1. Comparisons between offending and nonoffending youth are participant comparisons; this is a quasi-experiment.
2. We might assume that the children were independently assigned to each type of film, so this would be a true experiment.
3. It seems reasonable to assume that the pigeons were independently assigned to groups—a true experiment.

4. Gender is a participant variable; this is a quasi-experiment.

Conceptual Exercise 7B

This research problem is probably best studied in the field. It would be difficult to simulate rock concert conditions in a laboratory setting. And it is doubtful that the same kinds of behavior would occur in a laboratory as in the natural setting.

FAQ

Q1: Most psychology research is conducted on psychology students. Is that a problem?

A1: Using psychology students in research is a problem that affects the external validity of your study. To the degree that these participants are different from the population you want to generalize to, your study lacks external validity. Practically speaking, most researchers do not consider this a problem because their focus is on testing a null hypothesis. Whether experimental results conducted on university students will generalize to other groups is perhaps the topic of another experiment.

Q2: What is the difference between random assignment and random sampling?

A2: Random assignment is a procedure that is used to assign participants to treatment groups (or levels). The goal is to form groups of participants who are initially equivalent on measures of the DV. That is, if you were to measure the DV at the start of the experiment, the groups should not differ. Random selection is a procedure for selecting a sample from a population. The goal is to make your sample representative of the population. Random assignment affects the internal validity of your study,

and random selection affects the external validity.

Q3: Why is random assignment so important?

A3: Random assignment of participants to treatment groups (or levels of treatments) is crucial to the internal validity of the experiment. As mentioned in Answer 2, random assignment is used to ensure that the groups are equivalent on any variable that may influence the DV. For example, if your treatment and control groups differ on measures of the DV, it cannot be said that the difference was caused by the IV.

Q4: What is the difference between an experimental and a nonexperimental design?

A4: The main difference between an experimental and a nonexperimental design is control of the IV. In an experiment, the researcher has control over who receives the treatment. In a nonexperimental design, the participants have, in a sense, already been assigned to their group. For example, if you were conducting an experiment on the effects of alcohol on cognitive ability, you would assign sober participants to consume two, four, or six drinks and also assign some individuals to a nonalcoholic drink control group. Once everyone had consumed his or her assigned amount, you would administer your DV measure. This is an experiment because *you* have formed the groups. If you were to do this study using a nonexperimental design, you could simply enter a bar and ask people how many drinks they have had and then administer your DV measure. In the experiment, you could make causal statements about how the alcohol had influenced cognitive ability. In the nonexperimental design, you could make statements of relationship but not statements of causation. This is because you

did not assign the participants to the conditions; they assigned themselves. And it could be that the individuals differed in cognitive ability even before they began drinking (i.e., perhaps their differences in cognitive ability are driving them to drink excessively).

Q5: If the experiment is the cornerstone of scientific research, why use any other approach?

A5: Some research problems simply do not lend themselves to an experimental approach. Psychologists are interested in many variables that are inherent in the participants and therefore cannot be manipulated. Developmental psychologists, for example, often study variables such as gender and age. Such participant variables can only be studied with a quasi-experimental design. In addition, there may be ethical problems or practical reasons that cause the researcher to choose a nonexperimental approach.

Q6: Are treatment levels used in both experimental and nonexperimental designs?

A6: There are treatment groups in quasi-experimental designs; however, the researcher does not form them. When the researcher does not have control over the assignment of participants to groups, such a design is not considered a true experimental design, and strong statements of causal relationship cannot be made.

Q7: When doing field experiments, do you have to tell people they are being observed?

A7: Although you must have your research examined by an institutional review board, generally, you do not have to obtain consent when observing people in a public setting. However, you must be able to guarantee anonymity of those observed. In other words, if you use video, it must be done in such a way that no one can be identified.

Chapter Exercises

1. What are the main reasons for doing experiments?
2. What are the steps involved in conducting an experiment?
3. What is the difference between a controlled laboratory experiment and a field experiment?
4. Describe the advantages and disadvantages of controlled and field experiments.
5. How are participants assigned to groups in independent-groups designs, and what is the purpose of that method?
6. What does it mean when we say that there is a significant main effect?
7. What does it mean when we say that there is a significant interaction effect?

Chapter Projects

1. A social psychologist is interested in the effects of video games on children's hand-eye dexterity. Design a controlled experiment to investigate this research problem.
 - a. Specify the IV (operationalize).
 - b. Specify the DV (operationalize).
 - c. What is your research hypothesis?
 - d. Specify how you will select participants and assign them to conditions.
 - e. What is your statistical hypothesis, and how will you test it?
2. Design a field experiment to investigate the research problem described in Project 1.
 - a. Specify the IV (operationalize).
 - b. Specify the DV (operationalize).
 - c. What is your research hypothesis?
 - d. Specify how you will select participants and assign them to conditions.
 - e. What is your statistical hypothesis, and how will you test it?
3. Design an independent-groups experiment to evaluate the following conceptual hypothesis: Children diagnosed with attention deficit disorder are more distractible on group tasks than on individualized tasks.
 - a. Specify the IV (operationalize).
 - b. Specify the DV (operationalize).
 - c. What is your research hypothesis?
 - d. Specify how you will select participants and assign them to conditions.
 - e. What is your statistical hypothesis, and how will you test it?
4. You have conducted an experiment to determine how children diagnosed with attention deficit disorder perform on group and individualized tasks. In addition, the tasks are classified as difficult or easy. The children were independently assigned to each of the four conditions (difficult group, easy group, difficult individualized, and easy individualized). You have measured mean performance of the groups on task solution.

- What kind of design have you used?
- What are the IVs, and what is the DV?
- Using the data below, graph each main effect and the interaction effect using

group means. Describe what seems to have occurred. (Higher scores are better.)

Type of Task	Task Difficulty	
	Easy	Difficult
Individual	10.0	7.0
Group	9.0	2.0

Ancillaries

SAGE edge provides a personalized approach to help students accomplish their coursework goals in an easy-to-use learning environment. The site includes **flashcards** for key term practice, **learning objectives** to reinforce key materials, along with **open access media** for concept exploration. Visit the site at <https://edge.sagepub.com/rooney4e>.