CHAPTER EIGHT

Complex Designs

CHAPTER OUTLINE

Overview

In Chapters 6 and 7 we focused on the basic experimental designs that researchers use to study the effect of an independent variable. We described how an independent variable could be implemented with a separate group of participants in each condition (independent groups designs) or with each participant experiencing all the conditions (repeated measures designs). We limited our discussion to experiments involving only one independent variable because we wanted you to concentrate on the basics of experimental research. Experiments involving only one independent variable are not, however, the most common type of experiment in contemporary psychological research. Instead, researchers most often use **complex designs** in which two or more independent variables are studied simultaneously in one experiment.

Key Concept

Complex designs can also be called *factorial designs* because they involve factorial combination of independent variables. *Factorial combination* involves pairing each level of one independent variable with each level of a second independent variable. This makes it possible to determine the effect of each independent variable alone (*main effect*) and the effect of the independent variables in combination (*interaction effect*).

Complex designs may seem a bit complicated at this point, but the concepts will become clearer as you progress through this chapter. We begin with a review of the characteristics of experimental designs that can be used to investigate independent variables in a complex design. We then describe the procedures for producing, analyzing, and interpreting main effects and interaction effects. We introduce the analysis plans that are used for complex designs. We conclude the chapter by giving special attention to the interpretation of interaction effects in complex designs.

Describing Effects in a Complex Design

- Researchers use complex designs to study the effects of two or more independent variables in one experiment.
- In complex designs, each independent variable can be studied with an independent groups design or with a repeated measures design.
- The simplest complex design is a 2×2 design—two independent variables, each with two levels.
- The number of different conditions in a complex design can be determined by multiplying the number of levels for each independent variable (e.g., $2 \times 2 = 4$).
- More powerful and efficient complex designs can be created by including more levels of an independent variable or by including more independent variables in the design.

An experiment with a complex design has, by definition, more than one independent variable. Each independent variable in a complex design must be implemented using either an independent groups design or a repeated measures design according to the procedures described in Chapters 6 and 7. When a complex design has both an independent groups variable and a repeated measures variable, it is called a *mixed design.*

The simplest possible experiment involves one independent variable manipulated at two levels. Similarly, the simplest possible complex design experiment involves two independent variables, each with two levels. Complex designs are identified by specifying the number of levels of each of the independent variables in the experiment. A 2 × 2 (which is read "2 by 2") design, then, identifies the most basic complex design. Conceptually, there is an unlimited number of complex designs because any number of independent variables can be studied and each independent variable can have any number of levels. In practice, however, it is unusual to find experiments involving more than four or five independent variables, and two or three is more typical. Regardless of the number of independent variables, the number of conditions in a complex design can be determined by multiplying the number of levels of the independent variables. For example, if there are two independent variables with each having two levels (a 2 × 2 design), there are four conditions. In a 3 × 3 design there are two independent variables with three levels each, so there are nine conditions. In a 3 × 4 × 2 design there are three independent variables with three, four, and two levels, respectively, and a total of 24 conditions. The primary advantage of all complex designs is the opportunity they provide for identifying interactions between independent variables.

Understanding the 2 × 2 design lays a foundation for understanding complex designs. The 2 × 2 design barely scratches the surface, however, when it comes to tapping the potential of complex designs. Complex designs can be extended beyond the 2 × 2 design in one of two ways. Researchers can add levels to one or both of the independent variables in the design, yielding designs such as the 3 × 2, the 3 × 3, the 4 × 2, the 4 × 3, and so on. Researchers can also build on the 2 × 2 design by increasing the number of independent variables in the same experiment. The number of levels of each variable can range from 2 to some unspecified upper limit. The addition of a third or fourth independent variable yields designs such as the 2 × 2 × 2, the 3 × 3 × 3, the 2 × 2 × 4, the 2 × 3 × 3 × 2, and so on.

First we will illustrate main effects and interaction effects in the complex design by working through an example of a 2 × 2 design.

An Example of a 2 × 2 Design

The nature of main effects and interaction effects is essentially the same in all complex designs, but they can be seen most easily in a 2 × 2 design. For an example of this design we will draw from the rich literature in the field of psychology and law. There are few areas in the legal arena that have gone untouched by social scientists. Jury selection, the nature and credibility of eyewitnesses, race of the defendant, jury decision making, and attorney arguments are only some of the many topics investigated by researchers. Recall that in Chapter 6 we discussed a research study by Ceci (1993) on children's eyewitness testimony. In the study to be discussed here the researchers looked at variables that might lead to false confessions from suspects brought in for questioning.

Kassin, Goldstein, and Savitsky (2003) used a 2 × 2 design to investigate whether interrogators' expectations regarding a suspect's guilt or innocence influence the interrogation tactics they use. Kassin and his colleagues have conducted many studies to identify factors that lead to false confessions by innocent people. In the present study, Kassin et al. hypothesized that one potential reason for false confessions is that interrogators have a *confirmation bias* in which their initial beliefs about a suspect's guilt cause them to interrogate more aggressively, ask questions in a manner that presumes guilt, and cause suspects to behave defensively (which is interpreted as guilt). In general, this behavioral confirmation theory has three parts: (1) the perceiver forms a belief about a target person; (2) the perceiver behaves toward the person in ways that are consistent with the belief; and (3) the target person then responds in ways that support the perceiver's belief. Ultimately, in the criminal justice context the end result of this process can be a confession of guilt by an innocent person.

Kassin and his colleagues (2003) tested the behavioral confirmation theory in a clever experiment involving college student participants. Pairs of students participated as interrogators and suspects. "Interrogators" were asked to play the role of a detective trying to solve a case in which $100 was stolen from a locked cabinet. Importantly, the researchers manipulated the interrogator's expectations regarding the suspect's guilt. Half of the student interrogators were randomly assigned to the *guilty expectation* condition, in which the experimenter said that 4 out of every 5 suspects in the experiment actually committed the crime. Thus, these research participants were led to believe their chances of interrogating a guilty suspect were high (80% likelihood). In the *innocent expectation* condition, research participants were told their chance of interrogating a guilty suspect was low because only 1 out of 5 suspects was actually guilty (20%). This independent variable, *interrogator expectation*, was manipulated to initiate a confirmation bias among interrogators.

Other students played the role of suspect. Because suspects' behavior in an actual interrogation is influenced by their true guilt or innocence, Kassin et al. manipulated students' guilt or innocence using the independent variable, *suspect status*. In the *guilty* condition, students were asked to commit a mock theft in which they were instructed to enter a room, find a key hidden behind a VCR, use the key to open a cabinet, take $100, return the key, and leave with the $100. Students in the *innocent* condition were asked to approach the same room, knock on the door, wait for an answer (which did not occur), and then meet the experimenter. Half of the student-suspects were randomly assigned to the guilty role and half were assigned to the innocent role. All suspects were instructed to convince the interrogator of their innocence and to not confess. Interrogators were given the conflicting goals of trying to obtain a confession but also to determine whether the suspect was actually guilty or innocent. The interrogations were tape recorded.

Factorial combination of the two independent variables created four conditions in this 2×2 complex design:

1. Actual guilt/Guilty expectation
2. Actual guilt/Innocent expectation
3. Actual innocence/Guilty expectation
4. Actual innocence/Innocent expectation

Keep in mind that each group formed by the combination of variables represents a random group of participants. The design looks like this:

	Interrogator Expectation	
Suspect Status	Guilty	Innocent
Actual guilt	1	2
Actual innocence	3	4

Kassin et al. (2003) measured several dependent variables so that they could determine if there was converging evidence in support of the behavioral confirmation theory. For example, they measured dependent variables for the interrogators and suspects, and for new, additional participants who listened to the tape-recorded interrogations (much like potential jurors might hear). We will focus on three dependent variables from their experiment to illustrate main effects and interactions. Let's see what they found.

Main Effects and Interaction Effects

- The overall effect of each independent variable in a complex design is called a main effect and represents the differences among the average performance for each level of an independent variable collapsed across the levels of the other independent variable.
- An interaction effect between independent variables occurs when the effect of one independent variable differs depending on the levels of the second independent variable.

Key Concept

In any complex factorial design it is possible to test predictions regarding the overall effect of each independent variable in the experiment while ignoring the effect of the other independent variable(s). The overall effect of an independent variable in a complex design is called a **main effect.** We will examine two main effects Kassin and his colleagues observed in their experiment for two different dependent variables.

Prior to their interrogation of the suspect, student interrogators were given information about interrogation techniques, including a list of possible questions they could ask about the theft. Twelve questions were written as pairs (but presented randomly in the list). One question of the pair was written in such a way that the suspect's guilt was presumed (e.g., "How did you find the key that was hidden behind the VCR?") and the second question in the pair was written so as not to presume guilt (e.g., "Do you know anything about the key that was hidden behind the VCR?"). Student interrogators were asked to select six questions they might later want to ask. Thus, students could select from 0 to 6 questions that presumed guilt. Based on the behavioral confirmation theory, Kassin et al. predicted that interrogators in the guilty-expectation condition would select more guilt-presumptive questions than would interrogators

TABLE 8.1 A MAIN EFFECT OF INTERROGATOR EXPECTATION ON THE NUMBER OF GUILT-PRESUMPTIVE QUESTIONS

	Interrogator Expectation	
Suspect Status	Guilty	Innocent
Actual guilt	3.54	2.54
Actual innocence	3.70	2.66
Means for interrogator expectation	3.62	2.60

Hypothetical cell means based on Kassin et al. (2003).

in the innocent-expectation condition. Thus, they predicted a *main effect* of the interrogator-expectation independent variable.

The data for this dependent variable, number of guilt-presumptive questions selected, are presented in Table 8.1. The overall mean number of guilt-presumptive questions for participants in the guilty-expectation condition (3.62) is obtained by averaging the means of the actual-guilt and actual-innocence conditions for interrogators in the guilty-expectation condition: $(3.54 + 3.70)/2 = 3.62$. Similarly, the overall mean for the innocent-expectation condition is computed to be 2.60: $(2.54 + 2.66)/2 = 2.60$.[1] *The means for a main effect represent the overall performance at each level of a particular independent variable collapsed across (averaged over) the levels of the other independent variable*. In this case we collapsed (averaged) over the suspect status variable to obtain the means for the main effect of the interrogator expectation variable. The *main effect* of the interrogator-expectation variable is the difference between the means for the two levels of the variable $(3.62 - 2.60 = 1.02)$. In the Kassin et al. experiment, the main effect of the interrogator-expectation variable indicates that the overall number of guilt-presumptive questions selected was greater when interrogators expected a guilty suspect (3.62) than when they expected an innocent suspect (2.60). Inferential statistics tests confirmed that the main effect of interrogator expectation was statistically significant. This supported the researchers' hypothesis based on behavioral confirmation theory.

Let's now turn to a dependent variable for which there was a statistically significant main effect of the suspect-status independent variable. The researchers also coded the tape-recorded interviews to analyze the techniques used by the interrogators to obtain a confession. Student interrogators were given brief, written instructions regarding the powerful techniques police use to break down a suspect's resistance. Researchers counted the number of interrogator statements that reflected these persuasive techniques, such as building rapport, assertions of the suspect's guilt or disbelief in the suspect's statements, appeals

[1]The simple averaging of the values within each row and column to obtain the means for the main effects is possible only when there are equal numbers of participants contributing to each mean in the table. For procedures to calculate weighted means when the cells of the table involve different sample sizes, see Keppel (1991).

TABLE 8.2 A MAIN EFFECT OF SUSPECT STATUS ON THE NUMBER OF PERSUASIVE TECHNIQUES

	Interrogator Expectation		
Suspect Status	Guilty	Innocent	Means for Suspect Status
Actual guilt	7.71	6.59	7.15
Actual innocence	11.96	10.88	11.42

Hypothetical cell means based on Kassin et al. (2003).

to the suspect's self-interest or conscience, threats of punishment, promises of leniency, and presentation of false evidence.

The data for this dependent variable, number of persuasive techniques, are presented in Table 8.2. The overall mean number of persuasive techniques interrogators used when they interviewed suspects who were actually guilty was 7.15. This mean is computed by averaging across the two levels of the interrogator-expectation variable in the actual-guilt condition: (7.71 + 6.59)/2. The overall mean number of persuasive techniques used when interrogators interviewed a suspect who was actually innocent was 11.42, computed by averaging across the interrogator-expectation variable in the actual-innocence condition: (11.96 + 10.88)/2. The difference between these means (11.42 − 7.15 = 4.27) represents the main effect of the suspect-status independent variable. On average, interrogators used 4.27 more persuasive techniques when the suspect was actually *innocent* compared to guilty. Kassin and his colleagues were surprised by the finding that innocent suspects in both interrogator-expectation conditions were interrogated more aggressively than suspects who were actually guilty.

Finally, we can also examine data for which Kassin et al. observed an interaction effect between the interrogator-expectation and suspect-status independent variables. In a second phase of the experiment a new sample of students was asked to listen to the tape-recorded interrogation and to make judgments about the behavior of the interrogator and suspect. One question asked these students to rate on a 10-point scale how hard the interrogator worked to get a confession from the suspect, with higher numbers indicating greater effort. These data are presented in Table 8.3.

TABLE 8.3 AN INTERACTION EFFECT BETWEEN INTERROGATOR EXPECTATION AND SUSPECT STATUS ON EFFORT TO OBTAIN A CONFESSION

	Interrogator Expectation	
Suspect Status	Guilty	Innocent
Actual guilt	5.64	5.56
Actual innocence	7.17	5.85

Cell means provided by Dr. Saul Kassin.

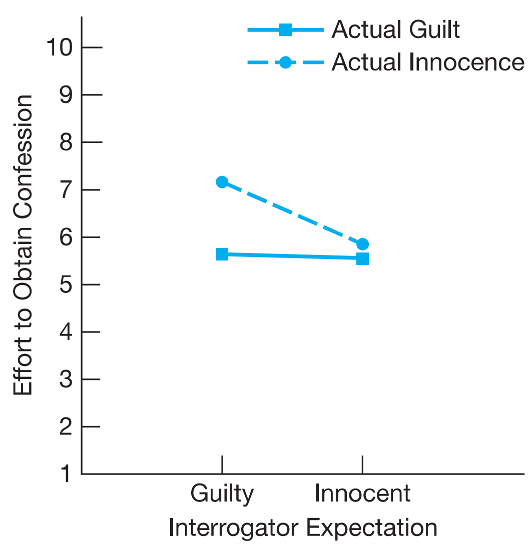
Key Concept

When two independent variables interact, we know that both variables together influence participants' performance on the dependent variable, in this case, ratings of the interrogators' effort to obtain a confession. Stated formally, an **interaction effect** occurs when the effect of one independent variable differs depending on the level of a second independent variable. To understand the interaction, examine the first row of Table 8.3. If only suspects who were actually guilty had been tested in the experiment, we would have concluded that the interrogators' expectations had *no effect* on effort ratings because the means for the guilty-expectation and innocent-expectation conditions are nearly identical. On the other hand, if only suspects who were actually innocent had been tested (second row of Table 8.3), we would have decided that interrogator expectations had a *large effect* on interrogators' efforts to obtain a confession.

An interaction effect is most easily seen when the means for the conditions are graphed. Figure 8.1 plots the four means found in Table 8.3. These results indicate that ratings of the interrogators' effort depend on whether the suspect is actually innocent or guilty *and* whether the interrogator expects the suspect to be guilty or innocent—that is, *both* independent variables are necessary to explain the effect. We describe the statistical analysis of interaction effects in complex designs in a later section, "Analysis of Complex Designs." For now, it is sufficient if you recognize that *an interaction effect occurs when the effect of one independent variable differs depending on the levels of a second independent variable*.

When one independent variable interacts with a second independent variable, the second independent variable must interact with the first one (that is, the order of the independent variables doesn't matter). For example, we described the interaction in Table 8.3 by stating that the effect of interrogators' expectations depends on the suspect's status. The reverse is also true; the effect of suspect status depends on the interrogators' expectations.

FIGURE 8.1 Graph illustrating the interaction effect between interrogator expectation and suspect status on effort to obtain a confession. (Data provided by Dr. Saul Kassin.)



STRETCHING EXERCISE I

In this exercise you are asked to examine Tables 8.1, 8.2, and 8.3 to answer the following questions.

1 (a) In Table 8.1, what are the means for the main effect of the suspect-status independent variable?
(b) How does the main effect of the suspect-status variable compare to the main effect of the interrogator-expectation variable for these data?
(c) Is an interaction effect likely present in these data?

2 (a) In Table 8.2, what are the means for the main effect of the interrogator-expectation independent variable?
(b) How does the main effect of the interrogator-expectation variable compare to the main effect of the suspect-status variable for these data?
(c) Is an interaction effect likely present in these data?

3 (a) In Table 8.3, what are the means for the main effect of the interrogator-expectation independent variable?
(b) What are the means for the main effect of the suspect-status independent variable?
(c) Kassin et al. (2003) observed these main effects to be statistically significant. Using the means you computed, describe the main effects of the interrogator-expectation and suspect-status variables in Table 8.3.

We are now in a position to describe the conclusions that Kassin et al. (2003) made based on their data analyses of all their data. Using behavioral confirmation theory, they hypothesized that interrogators' expectations of guilt would cause them to conduct an interrogation that would confirm their beliefs. Their results supported this hypothesis; overall, interrogators who suspected guilt conducted more aggressive interrogations. In turn, suspects in the guilty-expectation condition became more defensive and were perceived as guilty by the neutral observers. That the interrogators in the guilty-expectation condition were even more aggressive when trying to obtain a confession for suspects who were actually innocent demonstrates the power of their expectations of guilt and the power of the behavioral confirmation process. In the criminal justice context, police interrogations that are based on a preexisting bias of the suspect's guilt can trigger a biased chain of events that may lead to tragic conclusions, including false confessions by innocent people.

Describing Interaction Effects

- Evidence for interaction effects can be identified using descriptive statistics presented in graphs (e.g., nonparallel lines) or tables (subtraction method).
- The presence of an interaction effect is confirmed using inferential statistics.

How you choose to describe the results of an interaction effect depends on which aspect of the interaction effect you want to emphasize. For example, Kassin et al. (2003) emphasized the effect of the interrogation-expectation variable on innocent and guilty suspects to test their predictions based on behavioral confirmation theory. That is, the manipulation of interrogators' expectations of a suspect's guilt or innocence allowed them to test their predictions that interrogators would seek to confirm their expectations. By adding the second independent variable, Kassin et al. accomplished two things. First, the study more

STRETCHING EXERCISE II

In this exercise you will have the opportunity to practice identifying main effects and interaction effects in 2 × 2 complex designs using only descriptive statistics.

In the spirit of practice makes perfect, let us now turn our attention to the exercise we have prepared to help you learn to identify main effects and interaction effects. Your task is to identify main effects and interaction effects in each of six complex design experiments (A through F). In each table or graph in this box, you are to determine whether the effect of each independent variable differs depending on the level of the other independent variable. In other words, is there an interaction effect? After checking for the interaction effect, you can also check to see whether each independent variable produced an effect when collapsed across the other independent variable. That is, is there a main effect of one or both independent variables? The exercise will be most useful if you also practice translating the data presented in a table (Figure 8.2) into a graph and the data presented in graphs (Figures 8.3 and 8.4) into tables. The idea of the exercise is to become as comfortable as you can with the various ways of depicting the results of a complex design.

FIGURE 8.2 Mean number of correct responses as a function of task difficulty and anxiety level.

A.

		Task Difficulty	
		Easy	Hard
Anxiety Level	Low	3.3	3.3
	High	5.6	1.2

B.

		Task Difficulty	
		Easy	Hard
Anxiety Level	Low	6.2	3.1
	High	4.2	1.1

FIGURE 8.3 Mean number of aggressive responses as a function of type of media and content.

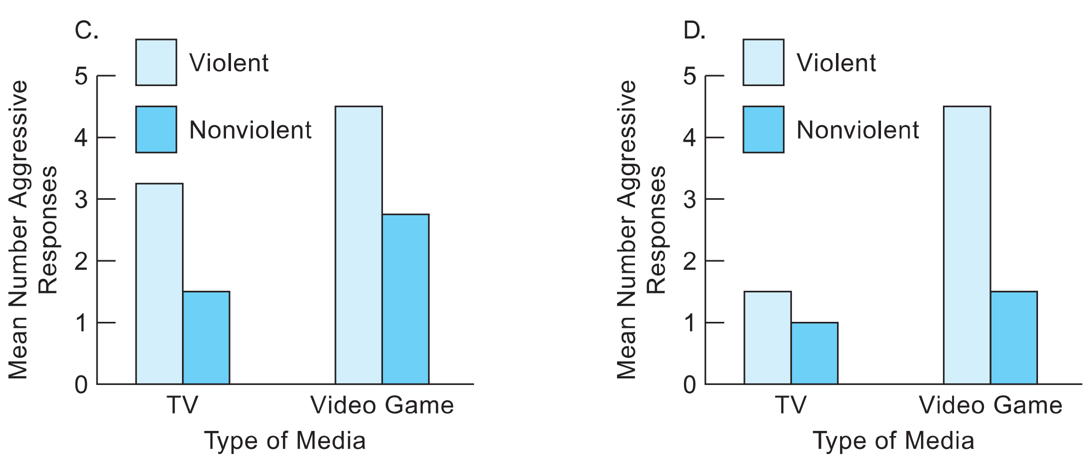
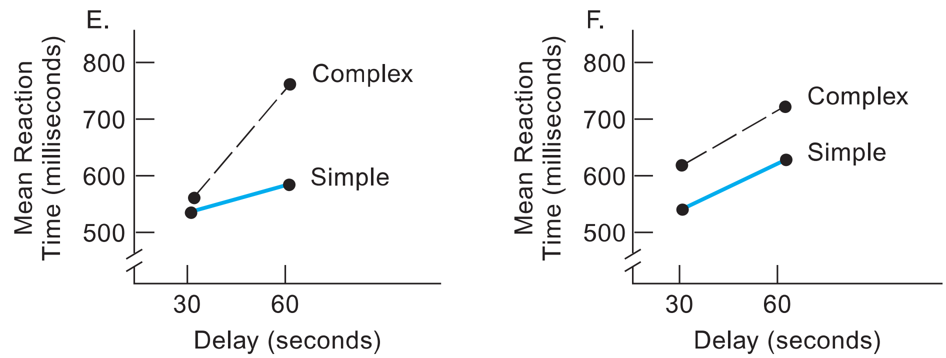


FIGURE 8.4 Mean reaction time as a function of delay and pattern complexity.



realistically conformed to real-world interrogations in which suspects are guilty or innocent; and second, they were able to demonstrate that interrogators who expect guilt work even harder to obtain a confession despite contrary evidence (e.g., the suspect's assertions of innocence). These findings also strongly indicate how the study of interaction effects in complex designs allows researchers to achieve greater understanding than is possible by doing experiments with only one independent variable.

There are three common ways to report a summary of the descriptive statistics in a complex design: tables, bar graphs, and line graphs. The procedures for preparing such tables and figures and the criteria for deciding which type of presentation to use are described in Chapter 13. In general, tables can be used for any complex design and are most useful when the exact values for each condition in the experiment need to be known. Bar graphs and line graphs, on the other hand, are especially useful for showing patterns of results without emphasizing the exact values. Line graphs are particularly useful for depicting the results of complex designs because an interaction effect can be seen so readily in a line graph. *Nonparallel lines in the graph suggest an interaction effect; parallel lines suggest no interaction effect*. See, for example, Figure 8.1.

When the results of a 2 × 2 design are summarized in a table, it is easiest to assess the presence or absence of an interaction effect by using the *subtraction method*. The subtraction method involves comparing the differences between the means in each row (or column) of the table. If the differences are different, an interaction effect is likely. In applying the subtraction method, it is essential that the differences be calculated in the same direction. For example, to use the subtraction method for the data reported in Table 8.3, you could subtract the mean ratings for the two levels of suspect status (actual guilt and actual innocence) for the guilty-expectation condition ($5.64 - 7.17 = -1.53$) and then do the same for the innocent-expectation condition ($5.56 - 5.85 = -0.29$). The sign of the obtained difference should also be carefully noted. The subtraction method shows you that these differences are different and, thus, an interaction effect between the two variables is likely. The subtraction method can be used only when one of the independent variables has two levels. For complex designs when both independent variables have three or more levels, graphs should be used to identify interaction effects.

Complex Designs with Three Independent Variables

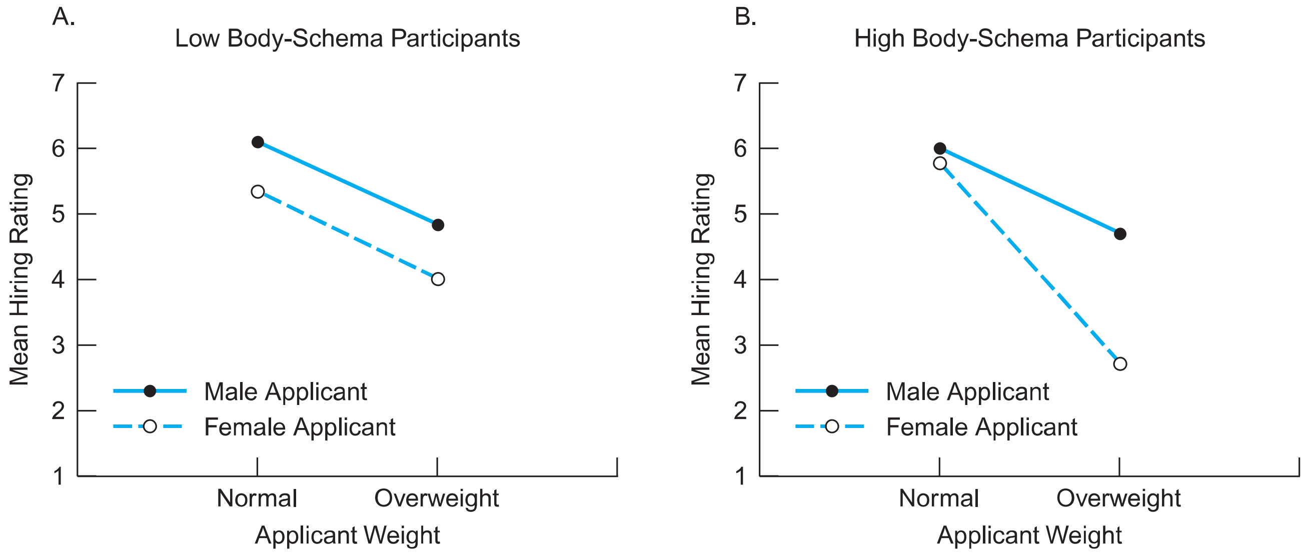
The power and complexity of complex designs increase substantially when the number of independent variables in the experiment increases from two to three. In the two-factor design there can be only one interaction effect, but in the three-factor design each independent variable can interact with each of the other two independent variables and all three independent variables can interact together. Thus, the change from a two-factor to a three-factor design introduces the possibility of obtaining four different interaction effects. If the three independent variables are symbolized as A, B, and C, the three-factor design allows a test of the main effects of A, B, and C; two-way interaction effects of A × B, A × C, B × C; and the three-way interaction effect of A × B × C. The efficiency of an experiment involving three independent variables is remarkable. An experiment investigating discrimination in the workplace will give you a sense of just how powerful complex designs can be.

Pingitore, Dugoni, Tindale, and Spring (1994) investigated possible discrimination against moderately obese people in a mock job interview. Participants in the experiment viewed videotapes of job interviews. In one of their experiments they used a 2 × 2 × 2 design. The first independent variable was the weight of the applicant (normal or overweight). The role of the applicant for the job in the videotapes was played by professional actors who were of normal weight. In the moderately obese conditions, the actors wore makeup and prostheses so that they appeared 20% heavier. The second independent variable in the experiment was the sex of the applicant (male or female). The third independent variable was participants' concern about their own body and the importance of body awareness to their self-concept (high or low). This variable was defined using a self-report measure of how participants viewed their body. A natural groups design was used to study this "body-schema variable." Participants were randomly assigned to evaluate male or female applicants who were normal weight or moderately obese (random groups designs). The dependent variable was the participants' rating on a 7-point scale of whether they would hire the applicant (1 = *definitely not hire* and 7 = *definitely hire*).

The results of the Pingitore et al. experiment for these three variables are shown in Figure 8.5. As you can see, displaying the means for a three-variable experiment requires a graph with more than one "panel." One panel of the figure shows the results for two variables at one level of the third variable, and the other panel shows results for the same two variables at the second level of the third independent variable.

As you are now familiar with main effects and simple (two-way) interaction effects, let us concentrate on understanding a three-factor or three-way interaction effect. As you can see in Figure 8.5, a two-way interaction effect of the applicant's weight and sex occurred only with participants who were high in concern about their own bodies. That is, those high on the body-schema variable

FIGURE 8.5 Illustration of an interaction effect for a 2 × 2 × 2 complex design.



(right panel of Figure 8.5) gave overweight female applicants especially low ratings but rated normal male and female applicants the same. Participants who were low on the body-schema variable (left panel of Figure 8.5), on the other hand, gave lower ratings to overweight applicants, but the difference between their ratings for male and female applicants was the same for both levels of the applicant weight variable.

One way to summarize the Pingitore et al. (1994) findings shown in Figure 8.5 is to say that the interaction effect of the independent variables of the applicants' weight and the applicants' sex depended upon the participants' body schema. We call this type of finding a three-way (or triple) interaction effect. As you can see, when we have a three-way interaction effect, all three independent variables must be taken into account when describing the results. In general, when there are two independent variables, an interaction effect occurs when the effect of one of the independent variables differs depending on the level of the second independent variable. *When there are three independent variables in a complex design, a three-way interaction effect occurs when the interaction of two of the independent variables differs depending on the level of the third independent variable*. The results shown in Figure 8.5 illustrate this well. The pattern of results for the first two independent variables (applicants' body weight and sex) differs depending on the level of the third variable (participants' body schema). By including the third independent variable of body-schema, Pingitore et al. provided a much better understanding of discrimination based on an applicant's weight than would have been the case had they included only the independent variables of sex and weight.

Analysis of Complex Designs

- In a complex design with two independent variables, inferential statistics are used to test three effects: the main effects for each independent variable and the interaction effect between the two independent variables.

- Descriptive statistics are needed to interpret the results of inferential statistics.
- How researchers interpret the results of a complex design differs depending on whether a statistically significant interaction effect is present or absent in the data.

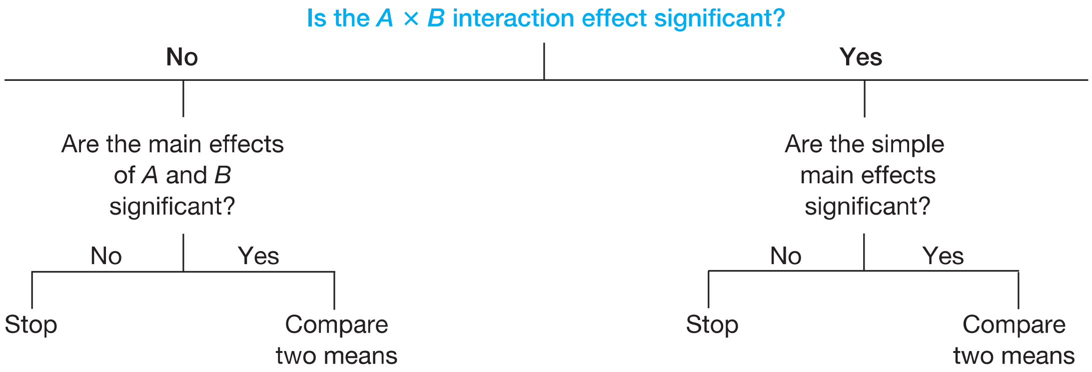
The analysis of complex designs builds on the logic used in the analysis of experiments with only one independent variable (see Chapters 6, 11, and 12). After checking the data for errors or outliers, the next step in data analysis is to describe the results using descriptive statistics such as mean, standard deviation, and measures of effect size. Inferential statistics such as null hypothesis testing and confidence intervals are then used to determine whether any of the effects are statistically reliable. On the basis of the descriptive and inferential statistics, researchers are able to make claims about what they have found.

Your task in the remaining section of this chapter is to understand data analysis as it is applied to complex designs, especially the manner in which an investigator interprets interaction effects and main effects. It may be helpful for you first to read the introduction that follows in this section, "Analysis of Complex Designs" and then to review the discussion of this topic in Chapter 12. The emphasis in both of these chapters is on the rationale and logic of these analyses, rather than on the nitty-gritty of computation. Fortunately, computers spare us the need to do the extensive calculations required of data produced in complex designs. On the other hand, computers cannot interpret the outcome of these calculations. That is where you come in. Go slowly; study this material carefully and be sure to examine the tables and figures that accompany the description in the text.

As you have come to understand, a complex design involving two variables has three potential *sources of systematic variation*. There are two potential main effects and a possible interaction effect. We describe the specific procedures for using null hypothesis testing (and the *F*-test) and confidence intervals to analyze complex designs in Chapter 12. A statistically significant effect in a complex design (as in any analysis) is an effect associated with a probability under the null hypothesis that is less than the accepted level of .05 (see Chapter 6). Inferential statistics tests are used in conjunction with descriptive statistics to determine whether an interaction effect has, in fact, occurred. After examining the data for an interaction effect, researchers may examine the data for the presence of main effects for each independent variable.

In a complex design, just as in an experiment with one independent variable, additional analyses may be needed to interpret the results. For example, a researcher might use confidence intervals to test differences between means. We illustrate such an approach in Chapter 12. The analysis plan for complex design experiments differs depending upon whether a statistically significant interaction effect is present in the experiment. Table 8.4 provides guidelines for interpreting a complex design experiment when an interaction effect does occur

TABLE 8.4 GUIDELINES FOR THE ANALYSIS OF A TWO-FACTOR EXPERIMENT



and when one does not. We will illustrate both paths in Table 8.4 by describing an experiment in which there is a statistically significant interaction effect present and then describing a study in which the interaction effect is not statistically significant.

Analysis Plan with an Interaction Effect

- If the analysis of a complex design reveals a statistically significant interaction effect, the source of the interaction effect is identified using simple main effects analyses and comparisons of two means.
- A simple main effect is the effect of one independent variable at one level of a second independent variable.

In order to understand the analysis of interaction effects within a complex design, we will examine a contemporary approach to understanding the effect of prejudice on individuals who are stigmatized. Social psychologists suggest that one effect of prejudice is that people who belong to stigmatized groups (e.g., ethnic minorities, gays and lesbians) develop belief systems about being devalued in society. With this "social-identity threat," stigmatized individuals develop expectations that cause them to be especially alert to cues in their environment that indicate they are viewed negatively (Kaiser, Vick, & Major, 2006). This attention to cues can occur at a *conscious* level, in which individuals are aware of their special attention to stigma cues. More recently, however, researchers have tested the extent to which social-identity threat causes people to be vigilant for potentially stigmatizing information without conscious awareness.

One method for examining nonconscious attention is the "emotional Stroop task." You may be familiar with the original version of the Stroop task in which participants are asked to name the color in which words are printed. The Stroop task was designed to show that reading is automatic (at least for adults). People find it impossible to ignore the printed words while naming the colors. This automatic-processing effect is demonstrated most dramatically in the condition in which color words are printed in a color other than the written word (e.g., "red" printed in blue ink). It takes participants longer to name colors in this "mismatch" condition because reading the word interferes with naming the color. Further studies show that this effect occurs even when the words are presented too quickly (e.g., 15 msec [milliseconds]) for participants to be consciously aware that a word was presented!

In the emotional Stroop task, the color words are replaced with content words that are particularly relevant to participants' concerns. For example, an experiment that examines nonconscious attention in people with phobias may use words such as "snake" and "spider." For phobic participants, identifying the color of these words takes longer than identifying words with neutral content, even when the words are presented subliminally (outside of conscious awareness).

Kaiser and her colleagues (2006) used the emotional Stroop task to investigate whether women with an expectation of being stigmatized through sexism would demonstrate greater nonconscious attention to sexist words compared to nonsexist words. They tested 35 women in a 2 × 3 complex design. The first manipulated independent variable was *social-identity* with two conditions, threat and safety, in a random groups design. Participants were led to believe that after completing the computer task, they would partner with a male participant (actually fictitious) to complete a group project. They were supplied with information about their partner so that they could get a sense of his personal characteristics. In the *identity-threat* condition, their partner held sexist views (e.g., strongly agreeing with statements such as "I could not work for a female boss because women can be overly emotional"). In the *identity-safety* condition, the partner was presented as nonsexist and strongly disagreed with sexist statements.

The second independent variable in their 2 × 3 design was *word type* with three levels: social-identity threatening, illness threatening, or nonthreatening. This variable was manipulated using a repeated measures design; thus, all participants were tested with all three word types in a completely counterbalanced order. The *social-identity threatening* words were sexist in content, such as *ho* and *hooters*. The *illness-threatening* words (e.g., *cancer, mono*) were included as a control condition to determine whether women in the identity-threat condition would pay attention to threatening words in general and not just social-identity threatening words. The *nonthreatening* words, also a control condition, described common household objects, such as *broom* and *curtains*. In one part of Kaiser et al.'s experiment, all three types of words were presented subliminally (15 msec) in different colors (red, yellow, blue, green), and participants' task was to identify the color. Tests showed that participants were unaware that words were presented. The dependent variable in this study was the response time for identifying the color (in milliseconds). This response-time measure assessed the amount of subliminal attention given to the different word types; longer response times indicate greater subliminal attention to the word and therefore a longer time to identify the color. The mean response times for each of the six conditions are presented in Table 8.5.

As Kaiser and her colleagues predicted, an interaction effect occurred between the two independent variables. Women in the identity-threat condition (first row of Table 8.5) took longest to name colors when social-identity threatening words were presented compared to illness-threatening and nonthreatening words. Longer response times to name the colors indicate that the women paid more subliminal attention to the words. Thus, women who expected to

TABLE 8.5 MEAN RESPONSE TIMES (IN MSEC) AS A FUNCTION OF SOCIAL IDENTITY AND WORD TYPE (SUBLIMINAL PRESENTATION)

	Word Type		
Social Identity Condition	Social-Identity Threatening	Illness Threatening	Non-Threatening
Threat	598.9	577.7	583.9
Safety	603.9	615.0	614.5

Data adapted from Kaiser et al. (2006).

interact with a sexist partner paid more subliminal attention to words that threatened their social identity. In contrast, women who anticipated interacting with a nonsexist man in the identity-safety condition (second row of Table 8.5) did not differ substantially in the attention given to the three different types of words. An interaction effect is present because the effect of the word-type variable differed depending on the level of the social-identity variable (threat, safety). Inferential statistics tests of these results using null hypothesis significance testing confirmed that the interaction effect was statistically significant.

Once an interaction effect is confirmed in the data, the specific source of the interaction is located using additional statistical tests. As outlined in Table 8.4, the specific tests for tracing the source of a significant interaction are called simple main effects and comparisons of two means (see Chapter 12).

Key Concept

A **simple main effect** is the effect of one independent variable at *one level* of a second independent variable. We can illustrate the use of simple main effects by returning to the results of the Kaiser et al. (2006) experiment. There are five simple main effects in Table 8.5: the effect of word type at each of the two levels of social identity and the effect of social identity at each of the three levels of word type. Kaiser et al. predicted that the subliminal attention effect (the difference between means for the three different word types) would occur for women in the identity-threat condition but not for women in the identity-safety condition. Therefore, they chose to test the simple main effects of word type at each level of the social-identity independent variable. They found, as predicted, that the simple main effect of word type was statistically significant in the identity-threat condition, but the simple main effect of word type was not statistically significant in the identity-safety condition.

When three or more means are tested in a simple main effect, as occurs for the word-type independent variable in Kaiser et al.'s experiment, comparisons of means tested two at a time can be done to identify the source of the simple main effect (see Chapter 12). First, no additional analyses are needed for the identity-safety condition because the simple main effect of word type was not statistically significant. The next step is to analyze the means more carefully for the identity-threat condition, where the simple main effect was statistically significant.

In their analyses of means considered two at a time, Kaiser and her colleagues noted both an expected and an unexpected effect for women in the identity-threat condition. As expected, mean response times were longer for social-identity threatening words than for illness-threatening words. Unexpectedly, mean response times did not differ when nonthreatening words were compared to either social-identity threatening words or to illness-threatening words. This raises an important question: Why did women allocate similar subliminal attention to nonthreatening words as they did to social-identity threatening words? Kaiser et al. reasoned that when women were expecting to interact with a sexist man, words describing household objects (e.g., *stove, broom, microwave*) in the nonthreatening condition may have been nonconsciously associated with sex-typed domestic tasks such as cooking and cleaning. According to Kaiser et al. (2006), "In retrospect, these nonthreatening words may not have provided the best comparison" (p. 336). Their interpretation of this unexpected finding illustrates how interpreting an experiment depends critically on how the experiment is done and how the data are analyzed.

Once an interaction effect has been thoroughly analyzed, researchers can also examine the main effects of each independent variable. However, the main effects are of much less interest when we know that an interaction effect occurred. For instance, the interaction effect in this experiment tells us that the subliminal attention given to different word types differs depending on the level of social-identity threat. Once we know this, we would not add much by learning whether, overall, women in the identity-safety condition had longer response times across all word types compared to women in the identity-threat condition. In the Kaiser et al. study, the main effects of the word-type and social-identity independent variables were not statistically significant. Nonetheless, there are experiments in which the interaction effect and the main effects are all of interest.

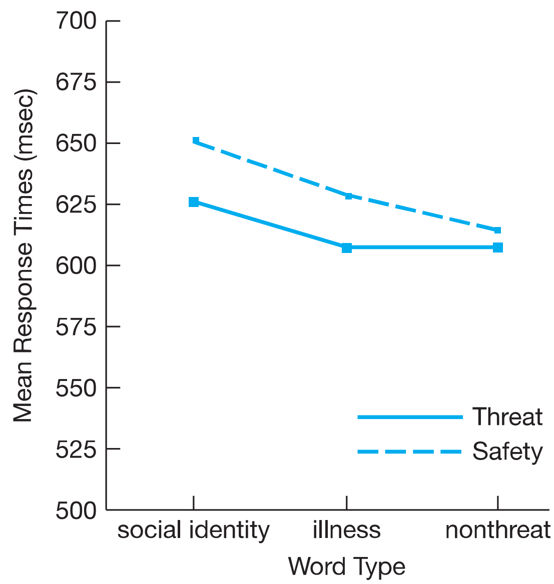
Analysis Plan with No Interaction Effect

- If the analysis of a complex design indicates the interaction effect between independent variables is not statistically significant, the next step in the analysis plan is to determine whether the main effects of the variables are statistically significant.
- The source of a statistically significant main effect can be specified more precisely by performing comparisons of two means or using confidence intervals to compare means two at a time.

We can use the results from a different part of the social-identity experiment conducted by Kaiser et al. (2006) to examine the analysis of a complex design when an interaction effect is *not* statistically significant. The results we just described were for words presented *subliminally*, that is, at a speed too fast (15 msec) for participants to detect the presence of the words. However, participants in this experiment also were tested with words presented at a *conscious* level. In the conscious-attention condition, women looked at the words on the screen until they responded by naming the color of the word.[2]

[2]The astute reader may see that the Kaiser et al. (2006) study is a 2 (social identity) × 3 (word type) × 2 (word presentation: subliminal, conscious) complex (mixed) design. The two levels of word presentation were manipulated using a repeated measures design. The 2 × 3 × 2 interaction among these independent variables was statistically significant. To further analyze the source of this three-way interaction, Kaiser et al. (2006) analyzed the 2 (social identity) × 3 (word type) interaction separately for subliminal presentation and conscious presentation. As described here, the 2 × 3 interaction was statistically significant for subliminal presentation but not for conscious presentation.

FIGURE 8.6 Results of a 2 × 3 complex design in which there was no interaction effect but there was a main effect. (Data provided by Dr. Cheryl R. Kaiser.)



The mean response times for the three word types (social-identity threatening, illness threatening, and nonthreatening) for the two different groups of women (identity-threat, identity-safety) are presented in Figure 8.6. The interaction effect, or, more accurately, the lack of an interaction effect, can be seen in the figure. Although the two lines in the figure are not perfectly parallel, the mean response times appear to decrease in both groups at approximately the same rate. Inferential statistics tests confirmed that the interaction effect was not statistically significant. The data shown in Figure 8.6 illustrate a general principle of data analysis: *The pattern of findings as shown by the descriptive statistics is not sufficient to decide whether an interaction effect is present in an experiment. Inferential statistics tests, such as the F-test; must be done to confirm whether the effects are statistically reliable.*

When the interaction effect is not statistically significant, the next step is to examine the main effects of each independent variable (see Table 8.4). The means for the Kaiser et al. conscious-awareness experiment are presented again in Table 8.6 to make it easier to determine the main effects. By collapsing

TABLE 8.6 MEAN RESPONSE TIMES (IN MSEC) AS A FUNCTION OF SOCIAL IDENTITY AND WORD TYPE (CONSCIOUS PRESENTATION)

	Word Type			
Social Identity Condition	Social-Identity Threatening	Illness Threatening	Non-Threatening	Means for Social Identity
Threat (n = 18)	625.9	607.4	607.5	613.6
Safety (n = 16)	650.6	629.0	614.5	631.4
Means for word type	637.5*	617.6*	610.8*	

Data provided by Dr. Cheryl R. Kaiser.
*Weighted means were calculated due to unequal sample sizes for the social-identity conditions.

(averaging) across the two social-identity conditions, we obtain the mean response times for each word type (i.e., for the main effect of the word-type variable). These means are 637.5 for the social-identity threatening words, 617.6 for the illness-threatening words, and 610.8 for the nonthreatening words. The main effect of word type was statistically significant. The source of a statistically significant main effect involving three or more means can be specified more precisely by comparing means two at a time (see Chapter 12). These comparisons can be done using *t*-tests or confidence intervals. Kaiser et al. found that, overall, women attended more (i.e., had longer response times) to the social-identity threatening cues ($M = 637.5$) than to both the illness-threatening cues ($M = 617.6$) and the nonthreatening cues ($M = 610.8$). There was no difference, however, between the latter two conditions. These results indicate that when consciously aware of the word types, women paid greater attention to words indicating a threat to their social identity.

We can also test for the main effect of the social-identity variable by using the means in Table 8.6. By collapsing across the word-type variable, we obtain the means for the identity-threat condition (613.6) and the identity-safety condition (631.4). The main effect of the social-identity variable was not statistically significant, indicating that, on average, response times were similar for women in the threat and safety conditions. That the two means appear to be different reinforces the need for statistical analyses to determine whether mean differences are reliable.

The analysis of Kaiser et al.'s social-identity experiment illustrates that much can be learned from a complex design even when there is no statistically significant interaction effect.

Interpreting Interaction Effects

Interaction Effects and Theory Testing

- Theories frequently predict that two or more independent variables interact to influence behavior; therefore, complex designs are needed to test theories.
- Tests of theories can sometimes produce contradictory findings. Interaction effects can be useful in resolving these contradictions.

Theories play a critical role in the scientific method. Complex designs greatly enhance researchers' ability to test theories because they can test for both main effects and interaction effects. For example, Kaiser et al. (2006) tested hypotheses about attention to prejudice cues in the environment based on social-identity theory. Prior research had demonstrated that when individuals' social identity is threatened, they are *consciously* aware of cues in their environment relating to potential prejudice. Kaiser et al. extended this research by testing the hypothesis that threatened individuals pay attention to prejudice cues *nonconsciously*, without awareness. Because they used a complex design, Kaiser et al.'s data provide evidence that women expecting to experience sexism, compared to women expecting a "safe" situation, paid greater subliminal attention to sexist words than to other words. Their data supported the social-identity theory of prejudice, in which "members of stigmatized groups develop belief systems about being devalued and that these expectations cause them to become especially alert or vigilant for signs of devaluation" (Kaiser et al., 2006, p. 332).

In addition, Kaiser et al. noted that theories of attentional processes state that attention is a limited resource. People who experience prejudice may allocate attention toward cues that threaten their social identity and therefore have less attentional resources available for other tasks. For example, students in a classroom setting who perceive possible prejudice may allocate their attention, both consciously and nonconsciously, to potential threats to their social identity, and this diverted attention could impair their classroom performance. Importantly, however, because Kaiser et al. manipulated the independent variable of social-identity threat with two levels, threat and safety, they were able to demonstrate that attentional resources are not diverted to potential threats when individuals believe they are safe from social-identity threats. This finding reinforces the importance of creating environments that are as free of prejudice as possible.

Psychological theories involving topics such as social identity and prejudice are often complex. In order to explain prejudice, for example, psychologists need to describe behavioral, cognitive, and emotional processes at individual, group, and societal levels. As you might imagine, experimental tests of complex theories can lead to contradictory findings. For example, consider a hypothetical example in which a study of prejudice shows that members of a devalued group do *not* experience heightened nonconcious attention to social-identity threats. How would this seemingly contradictory finding be incorporated into a theory of prejudice which states that stigmatized individuals attend to potential threats to their identity? As data from the Kaiser et al. experiment suggest, one interpretation of this finding might involve the independent variable of social-identity condition, threat or safety. The contradictory finding could be interpreted by suggesting that participants in the hypothetical study of prejudice felt safe from social-identity threats and therefore did not allocate attention to potential sources of devaluation.

A common approach to resolving contradictory findings is to include in the research design independent variables that address potential sources of contradictory findings (for example, by including threat and safety conditions in the design). More generally, complex designs can be extremely useful in tracking down the reasons for seemingly contradictory findings when theories are tested. The process can be a painstaking one, but it can also be very worthwhile.

Interaction Effects and External Validity

- When no interaction effect occurs in a complex design, the effects of each independent variable can be generalized across the levels of the other independent variable; thus, external validity of the independent variables increases.
- The presence of an interaction effect identifies boundaries for the external validity of a finding by specifying the conditions in which an effect of an independent variable occurs.

In Chapter 6 we discussed at some length the procedures for establishing the external validity of a research finding when an experiment involves only one independent variable. We described how partial replications could be done to establish external validity—that is, the extent to which research findings may be generalized. We also discussed how field experiments allow researchers to examine independent variables in real-world settings. We can now examine the role of complex designs in establishing the external validity of a finding. The presence or absence of an interaction effect is critical in determining the external validity of the findings in a complex design.

When no interaction effect occurs in a complex design, we know that the effects of each independent variable can be generalized across the levels of the other independent variable. For instance, consider again the findings from Kassin et al.'s (2003) study on interrogators' expectations when interrogating a suspect. They found that when interrogators expected the suspect to be guilty, they selected more guilt-presumptive questions than when interrogators expected the suspect to be innocent, regardless of whether the suspect was actually guilty or innocent. That is, there was no interaction effect between the interrogator-expectation variable and the suspect-status variable. Thus, interrogators' selection of guilt-presumptive questions when they expect guilt can be generalized across situations in which the suspect is actually guilty or innocent.

Of course, we cannot generalize our findings beyond the boundaries or conditions that were included in the experiment. For example, the absence of an interaction effect between interrogator expectations and suspect status does not allow us to conclude that the selection of guilt-presumptive questions would be similar if other groups were tested, such as law enforcement officials. Similarly, we do not know whether the same effects would occur if other manipulations of interrogators' expectations were used. We also must remember that not finding a statistically significant interaction effect does not necessarily mean that an interaction effect is not really present; we may not have performed an experiment with sufficient sensitivity to detect it.

As we have seen, the absence of an interaction effect increases the external validity of the effects of each independent variable in the experiment. Perhaps more important, the *presence* of an interaction effect identifies boundaries for the external validity of a finding. For example, Kassin et al. (2003) also found that interrogators who expected the suspect to be guilty, rather than innocent, applied greater pressure to obtain a confession on suspects who were actually innocent compared to those who were guilty. This interaction effect clearly sets limits on the external validity of the effect of interrogators' expectations on pressure to obtain a confession. Given this finding, the best way to respond to someone's query regarding the general effect of interrogators' expectations on their effort to obtain a confession is to say, "It depends." In this case, it depends on whether the suspect is actually guilty or innocent. The presence of the interaction effect sets boundaries for the external validity, but the interaction effect also specifies what those boundaries are.

The possibility of interaction effects among independent variables should lead us to be cautious about saying that an independent variable does not have an effect on behavior. Independent variables that influence behavior are called

Key Concept

relevant independent variables. In general, a **relevant independent variable** is one that influences behavior directly (results in a main effect) or produces an interaction effect when studied in combination with a second independent variable. Distinguishing between factors that affect behavior and those that do not is essential for developing adequate theories to explain behavior and for designing effective interventions to deal with problems in applied settings such as schools, hospitals, and factories (see Chapters 9 and 10).

There are several reasons why we should be cautious about identifying an independent variable as *irrelevant*. First, if an independent variable is shown to have no effect in an experiment, we cannot assume that this variable wouldn't have an effect if different levels of the independent variable had been tested. Second, if an independent variable has no effect in a single-factor experiment, this doesn't mean that it won't interact with another independent variable when used in a complex design. Third, if an independent variable does not have an effect in an experiment, it may be that an effect could have been seen with different dependent variables. Fourth, the absence of a statistically significant effect may or may not mean that the effect is not present. Minimally, we would want to consider the sensitivity of our experiment and the power of our statistical analysis before deciding that we have identified an irrelevant variable. (See Chapter 12 for a discussion of the power of a statistical analysis.) For now, it is best if you avoid being dogmatic about identifying any independent variable as not having any effect.

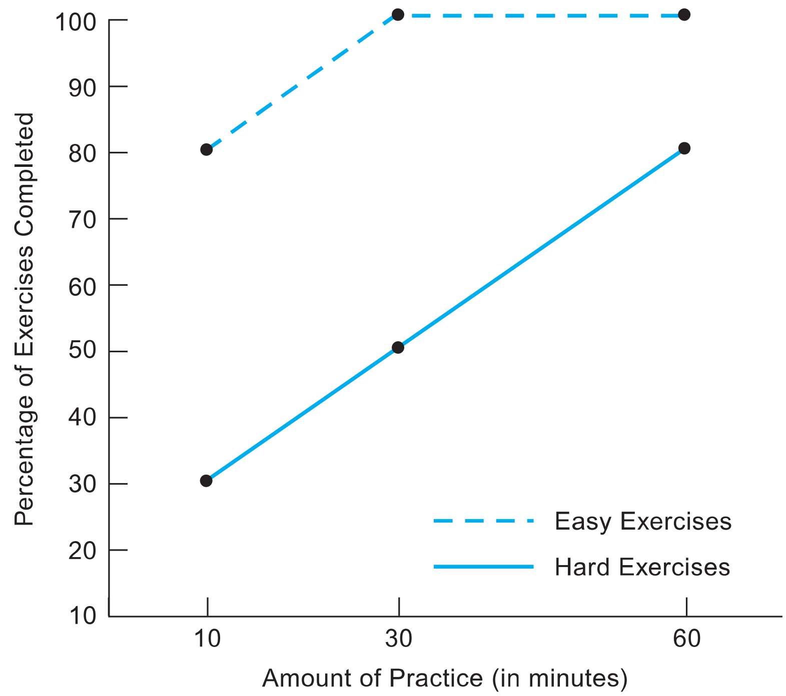
Interaction Effects and Ceiling and Floor Effects

- When participants' performance reaches a maximum (ceiling) or a minimum (floor) in one or more conditions of an experiment, results for an interaction effect are uninterpretable.

Consider the results of a 3 × 2 experiment investigating the effects of increasing amounts of practice on performance during a physical-fitness test. There were six groups of participants in this plausible but hypothetical experiment. Participants were first given 10, 30, or 60 minutes to practice, doing either easy or hard exercises. Then they took a fitness test using easy or hard exercises (the same they had practiced). The dependent variable was the percentage of exercises that each participant was able to complete in a 15-minute test period. Results of the experiment are presented in Figure 8.7.

The pattern of results in Figure 8.7 looks like a classic interaction effect; the effect of amount of practice time differed for the easy and hard exercises. Increasing practice time improved test performance for the hard exercises, but performance leveled off after 30 minutes of practice with the easy exercises. If a standard analysis was applied to these data, the interaction effect would very likely be statistically significant. Unfortunately, this interaction effect would be essentially uninterpretable. For those groups given practice with the easy exercises, performance reached the maximum level after 30 minutes of practice, so no improvement beyond this point could be shown in the 60-minute group. Even if the participants given 60 minutes of practice had further benefited from the extra practice, the experimenter could not measure this improvement on the chosen dependent variable.

FIGURE 8.7 Illustration of a ceiling effect.



Key Concepts

The preceding experiment illustrates the general measurement problem referred to as a ceiling effect. Whenever performance reaches a maximum in any condition of an experiment, there is danger of a **ceiling effect.** The corresponding name given to this problem when performance reaches a minimum (e.g., zero errors on a test) is a **floor effect.** Researchers can avoid ceiling and floor effects by selecting dependent variables that allow ample "room" for performance differences to be measured across conditions. For example, in the fitness experiment it would have been better to test participants with a greater number of exercises than anyone could be expected to complete in the time allotted for the test. The mean number of exercises completed in each condition could then be used to assess the effects of the two independent variables without the danger of a ceiling effect. It is important to note that ceiling effects also can pose a problem in experiments that don't involve a complex design. If the fitness experiment had included only the easy exercises, there would still be a ceiling effect in the experiment.

Interaction Effects and the Natural Groups Design

- Researchers use complex designs to make causal inferences about natural groups variables when they test a theory for why natural groups differ.
- Three steps for making a causal inference involving a natural groups variable are to state a theory for why group differences exist, manipulate an independent variable that should demonstrate the theorized process, and test whether an interaction effect occurs between the manipulated independent variable and natural groups variable.

The natural groups design, described briefly in Chapter 6, is one of the most popular research designs in psychology. Groups of people are formed by selecting individuals who differ on some characteristic such as gender, age, introversion–extraversion, or aggressiveness, to name just a few individual differences variables. Researchers then look for systematic relationships between these individual differences variables and other aspects of behavior. The natural groups design is an effective one for establishing correlations between individuals' characteristics and their performance. As we also described in Chapter 6, however, the natural groups design is perhaps the most challenging design when it comes to drawing conclusions about the causes of behavior.

The difficulty in interpreting the natural groups design arises when we try to conclude that differences in performance are *caused* by the characteristics of the people we used to define the groups. For instance, consider an experiment in which participants are selected because of their musical training. One group of participants includes people with 10 or more years of formal musical training, and one group includes people with no formal training. Both groups are tested on their ability to remember the musical notation for simple 10-note melodies. Not surprisingly, the results of these tests show that those with musical training perform far better than those without such training.

We can conclude on the basis of these results that memory for simple melodies varies with (is correlated with) amount of musical training. But we cannot conclude that musical training *causes* superior memory performance. Why not? There are probably many additional ways in which people with 10 years of musical training differ from those without such training. The groups may differ in amount and type of general education, family background, socioeconomic status, and amount and type of experience they have had listening to music. Also, those with musical training may have generally better memories than those without such training, and their superior memory for simple melodies may reflect this general memory ability. Finally, those who sought out musical training may have done so because they had a special aptitude for music. Accordingly, they might have done better on the memory task even if they had not had any musical training. In short, there are many possible causes other than individual differences in musical training for the difference in memory performance that was observed.

There is a potential solution to the problem of drawing causal inferences based on the natural groups design (Underwood & Shaughnessy, 1975). The key to this solution is to develop a theory regarding the critical individual difference variable. For example, Halpern and Bower (1982) were interested in how memory for musical notation differs between musicians and nonmusicians. Halpern and Bower developed a theory of how musical training would influence the cognitive processing of musical notation by those who had such training. Their theory was based on a memory concept called "chunking." You can get some sense of the memory advantage provided by chunking if you imagine trying to memorize the following strings of 15 letters: HBOFBICNNUSAWWW. Chunking helps memory by changing the same string of letters to a series of five more easily remembered chunks: HBO-FBI-CNN-USA-WWW.

Halpern and Bower theorized that musical training led musicians to "chunk" musical notation into meaningful musical units, thereby reducing the amount of information they needed to remember in order to reproduce the notation for a simple melody. Furthermore, if this process were responsible for the difference between the memory performance of musicians and nonmusicians, then the difference between musicians and nonmusicians should be greater for melodies with good musical structure than for melodies with poor musical structure. Halpern and Bower manipulated the independent variable of musical structure to test their theory. To do this, they used three different types of melodies to test their groups of musicians and nonmusicians. They prepared sets of simple melodies whose notations had similar visual structures but that were good, bad, or random in musical structure.

The critical test in Halpern and Bower's experiment was whether they would obtain an interaction effect between the two independent variables: musical training and type of melodies. Specifically, they expected that the difference in memory performance between musicians and nonmusicians would be largest for the melodies exhibiting good structure, next largest for the melodies exhibiting bad structure, and smallest for the random melodies. The results of Halpern and Bower's experiment conformed exactly to their predictions.

The obtained interaction effect allowed Halpern and Bower to rule out many alternative hypotheses for the difference in memory performance between musicians and nonmusicians. Such characteristics as amount and type of general education, socioeconomic status, family background, and good memory ability are not likely to explain why there is a systematic relationship between the structure of the melodies and the size of the difference in memory performance between musicians and nonmusicians. These potential alternative hypotheses cannot explain why there was little difference in the two groups' memory performance for random melodies. The interaction effect makes such simple correlational explanations much less plausible.

There are several steps that the investigator must take in carrying out the general procedure for drawing causal inferences based on the natural groups design.

Step 1: Develop a Theory The first step is to develop a theory explaining why a difference should occur in the performance of groups that have been differentiated on the basis of an individual differences variable. For example, Halpern and Bower theorized that musicians and nonmusicians differed in musical performance because of the way that these groups cognitively organize ("chunk") melodies.

Step 2: Identify a Relevant Variable to Manipulate The second step is to select an independent variable that can be manipulated and that is presumed to influence the likelihood that this theoretical process will occur. Halpern and Bower suggested that type of musical structure was a variable associated with ease of chunking.

Step 3: Test for an Interaction The most critical aspect of the recommended approach is to strive to produce an interaction effect between the manipulated variable and the individual differences variable. Thus, the relevant manipulated independent variable is applied to both natural groups. Halpern and Bower sought an interaction effect between the individual differences variable (musician vs. nonmusician) and the manipulated variable (type of musical structure) in a 2×3 complex design. The approach can be strengthened even further by testing predictions of interaction effects of three independent variables: two manipulated independent variables and the individual differences variable (see, for example, Anderson & Revelle, 1982).

SUMMARY

A complex design is one in which two or more independent variables are studied in the same experiment. A complex design involving two independent variables allows researchers to determine the overall effect of each independent variable (the main effect of each variable). More important, complex designs can be used to reveal the interaction effect between independent variables. Interaction effects occur when the effect of each independent variable depends on the level of the other independent variable.

The simplest possible complex design is the 2×2 design, in which two independent variables are both studied at two levels. The number of conditions in a factorial design is equal to the product of the levels of the independent variables (e.g., $2 \times 3 = 6$). Complex designs beyond the 2×2 can be even more useful for understanding behavior. Additional levels of one or both of the independent variables can be added to yield designs such as the 3×2, the 3×3, the 4×2, the 4×3, and so on. Additional independent variables can also be included to yield designs such as the $2 \times 2 \times 2$, the $2 \times 3 \times 3$, and so on. Experiments involving three independent variables are remarkably efficient. They allow researchers to determine the main effects of each of the three variables, the three two-way interaction effects, and the simultaneous interaction effect of all three variables.

When two independent variables are studied in a complex design, three potential sources of systematic variation can be interpreted. Each independent variable can produce a statistically significant main effect, and the two independent variables can combine to produce a statistically significant interaction effect. Interaction effects can be initially identified by using the subtraction method when the descriptive statistics are reported in a table, or by the presence of nonparallel lines when the results appear in a line graph. If the interaction effect does prove to be statistically significant, we can analyze the results further by examining simple main effects and, if necessary, comparisons of means considered two at a time. When no interaction effect arises, we examine the main effects of each independent variable, and we can use comparisons of two means or confidence intervals when necessary.

Complex designs play a critical role in the testing of predictions derived from psychological theories. Complex designs are also essential to resolve contradictions that arise when theories are tested. When a complex design is used and no interaction effect occurs, we know that the effects of each independent variable can be generalized across the levels of the other independent variable(s). When an interaction effect does occur, however, boundaries on the external validity of a finding can be clearly specified. The possibility of interaction effects requires that we expand the definition of a relevant independent variable to include those that influence behavior directly (produce main effects) and those that produce an interaction effect when studied in combination with another independent variable. Interaction effects that may arise because of measurement problems such as ceiling or floor effects must not be confused with interaction effects that reflect the true combined effect of two independent variables. Interaction effects can also be most helpful in solving the problem of drawing causal inferences based on the natural groups design.

Key Concepts

complex designs 250
main effect 253
interaction effect 256
simple main effect 265
relevant independent variable 271
ceiling and floor effects 272

Review Questions

1 Identify the number of independent variables, the number of levels for each independent variable, and the total number of conditions for each of the following examples of complex design experiments: (a) 2×3 (b) 3×3 (c) $2 \times 2 \times 3$ (d) 4×3.
2 Identify the conditions in a complex design when the following independent variables are factorially combined: (1) type of task with three levels (visual, auditory, tactile) and (2) group of children tested with two levels (normal, developmentally delayed).
3 Use the Kassin et al. results in Table 8.3 for interrogators' efforts to obtain a confession to show there are two possible ways to describe the interaction effect.
4 Describe how you would use the subtraction method to decide whether an interaction effect was present in a table showing the results of a 2×2 complex design.
5 Describe the pattern in a line graph that indicates the presence of an interaction effect in a complex design.
6 Outline the steps in the analysis plan for a complex design with two independent variables when there is an interaction effect and when there is not an interaction effect.
7 Use an example to illustrate how a complex design can be used to test predictions derived from a psychological theory.
8 How is the external validity of the findings in a complex design influenced by the presence or absence of an interaction effect?
9 Explain why researchers should be cautious about saying that an independent variable does not have an effect on behavior.
10 Describe the pattern of descriptive statistics that would indicate a ceiling (or floor) effect may be present in a data set, and describe how this pattern of data may affect the interpretation of inferential statistics (e.g., F-test) for these data.
11 Explain how interaction effects in a complex design can be used as part of the solution to the problem of drawing causal inferences on the basis of the natural groups design.

CHALLENGE QUESTIONS

1 Consider an experiment in which two independent variables have been manipulated. Variable *A* has been manipulated at three levels, and Variable *B* has been manipulated at two levels.

A Draw a graph showing a main effect of Variable *B*, no main effect of Variable *A*, and no interaction effect between the two variables.

B Draw a graph showing no main effect of Variable *A*, no main effect of Variable *B*, but an interaction effect between the two variables.

C Draw a graph showing a main effect of Variable *A*, a main effect of Variable *B*, and no interaction effect between the *A* and *B* variables.

2 A researcher has used a complex design to study the effects of training (untrained and trained) and problem difficulty (easy and hard) on participants' problem-solving ability. The researcher tested a total of 80 participants, with 20 randomly assigned to each of the four groups resulting from the factorial combination of the two independent variables. The data presented below represent the mean percentage of the problems that participants solved in each of the four conditions.

	Training	
Problem Difficulty	Untrained	Trained
Easy	90	95
Hard	30	60

A Is there evidence of a possible interaction effect in this experiment?

B What aspect of the results of this experiment would lead you to be hesitant to interpret an interaction effect if one were present in this experiment?

C How could the researcher modify the experiment so as to be able to interpret an interaction effect if it should occur?

3 A psychologist is interested in whether older people suffer a deficit with respect to their reaction time in processing complex visual patterns. Fifty 65-year-old people and 50 college-age young adults volunteer to participate in the experiment. The participants are tested using an embedded figures test. The psychologist presents a simple figure to each participant followed immediately by a complex pattern that contains the simple figure. The participant must indicate as quickly as possible the location of the simple figure in the complex pattern. Participants are timed from the onset of the complex pattern until they locate the simple pattern. As the psychologist had expected, the mean reaction times for the older adults were markedly longer than those for the young adults. By any standard the results were statistically significant.

A The psychologist claims based on these results that the differences in reaction times in this experiment were caused by a deficit in the older adults' ability to process complex information. You recognize that a complex design experiment would need to be done before he could conclude that older adults suffered a deficit in their processing of *complex* visual patterns. What additional reaction-time test could the psychologist give to both groups in order to make his experiment into a complex design? Describe an outcome of the complex design experiment that would support the claim that older adults suffer a deficit in processing complex information and another outcome that would lead you to question the claim.

B Recognizing that his original study is flawed, the psychologist tries to use post hoc (after the fact) matching to try to equate his two groups. He decides to match on general health (i.e., the better your general health, the faster your reaction time). Although he cannot get an exact matching across groups, he does find that when he looks only at the 15 healthiest older adults, their reaction times are only slightly longer than the mean for the young adults. Explain how this outcome would change the psychologist's conclusion concerning the effect of age on reaction time. Could the psychologist reach the general conclusion that older adults do not suffer a deficit in reaction time in this task? Why or why not?

Answer to Stretching Exercise I

1 (a) Actual guilt: $M = 3.04$, Actual innocence: $M = 3.18$

(b) The difference between the means for the suspect-status independent variable is 0.14, which is a very small difference compared to the mean difference observed for the statistically significant effect of interrogator expectation on the number of presumptive questions ($3.62 - 2.60 = 1.02$).

(c) Using the subtraction method, the difference between the actual-guilt and actual-innocent conditions in the guilty-expectation condition is −0.16 (3.54 − 3.70). In the innocent-expectation condition, this difference is −0.12 (2.54 − 2.66). Because these differences are very similar, an interaction effect is unlikely.

2 (a) Expect guilty: $M = 9.84$, Expect innocent: $M = 8.74$

(b) The difference between the means for the expect-guilty and expect-innocent conditions is 1.1 (i.e., approximately 1 more persuasive technique in the expect-guilty condition than in the expect-innocent condition). In contrast, for the statistically significant main effect of the suspect-status independent variable, the difference in the number of persuasive techniques used between actual-guilt and actual-innocent conditions is 4.27 (11.42 − 7.15).

(c) An interaction effect is unlikely. Using the subtraction method, the difference between the actual-guilt and actual-innocent conditions in the guilty-expectation condition (7.71 − 11.96 = −4.25) is very similar to the computed value for the innocent-expectation condition (6.59 − 10.88 = −4.29).

3 (a) Expect guilty: $M = 6.40$, Expect innocent: $M = 5.70$

(b) Actual guilt: $M = 5.60$, Actual innocence: $M = 6.51$

(c) The statistically significant main effect of the interrogator-expectation variable indicates that effort to obtain a confession (the dependent variable) was higher in the expect-guilty condition ($M = 6.40$) than in the expect-innocent condition ($M = 5.70$).

The statistically significant main effect of the suspect-status variable indicates that effort to obtain a confession was higher in the actual-innocence condition ($M = 6.51$) than in the actual-guilt condition ($M = 5.60$).

Answer to Stretching Exercise II

A interaction effect, main effect of the task difficulty

B no interaction effect, main effects of task difficulty and anxiety level

C no interaction effect, main effects of type of media and content

D interaction effect, main effects of type of media and content

E interaction effect, main effects of delay and pattern complexity (additional statistical analyses are needed to test these effects)

F no interaction effect, main effects of delay and pattern complexity

Answer to Challenge Question 1