MODULE 14

Advanced Experimental Designs

LEARNING OBJECTIVES

- Explain what additional information can be gained by using designs with more than two levels of an independent variable.
- Explain factorial notation and the advantages of factorial designs.
- Identify main effects and interaction effects based on looking at graphs.
- Draw graphs for factorial designs based on matrices of means.

The experiments described in the previous two modules typically involve manipulating one independent variable with only two levels, either a control group and an experimental group or two experimental groups. In this module we discuss experimental designs that involve one independent variable with more than two levels. Examining more levels of an independent variable allows us to address more complicated and interesting questions. Often experiments begin as two-group designs and then develop into more complex designs as the questions asked become more elaborate and sophisticated. The same design principles presented in Modules 12 and 13 apply to these more complex designs; that is, we still need to be concerned about control, internal validity, and external validity.

factorial design: A design with more than one independent variable.

In addition, we discuss designs with more than one independent variable. These are usually referred to as **factorial designs**, indicating that more than one *factor*, or independent variable, is manipulated in the study (an independent variable is often referred to as a factor). We discuss the advantages of such designs over simpler designs. Further, we explain how to interpret the findings (called main effects and interaction effects) from such designs.

USING DESIGNS WITH MORE THAN TWO LEVELS OF AN INDEPENDENT VARIABLE

Researchers may decide to use a design with more than two levels of an independent variable for three reasons. First, it allows them to compare multiple treatments. Second, it enables them to compare multiple treatments with no treatment (the control group). Third, more complex designs allow researchers to compare a placebo group with control and experimental groups (Mitchell & Jolley, 2004).

Comparing More Than Two Kinds of Treatment in One Study

To illustrate this advantage of more complex experimental designs, imagine that we want to compare the effects of various types of rehearsal on memory. We have participants study a list of 10 words using either rote rehearsal (repetition) or some form of elaborative rehearsal. Additionally, we specify the type of elaborative rehearsal to be used in the different experimental groups. Group 1 (the control group) uses rote rehearsal, group 2 uses an imagery mnemonic technique, and group 3 uses a story mnemonic device.

Notice that we do not simply conduct three studies or comparisons of group 1 to group 2, group 2 to group 3, and group 1 to group 3. Doing so is not recommended for several reasons. One reason has to do with the statistical analysis of the data. Although statistical concepts appear later in this text, you may be familiar with some statistical concepts and see the problem with making multiple comparisons. Each comparison must be statistically analyzed on its own, and this process can lead to an erroneous conclusion. The most likely error when making multiple comparisons is that they can lead us to conclude that there is an effect when in reality there is not.

Another advantage of comparing more than two kinds of treatment in one experiment is that it reduces both the number of experiments conducted and the number of participants needed. Once again, refer to the three-group memory experiment. If we do one comparison with three groups, we can conduct only one experiment, and we need participants for only three groups. However, if we conduct three comparisons, each with two groups, then we need to perform three experiments, and we need participants for six groups or conditions.

Comparing Two or More Kinds of Treatment with the Control Group (No Treatment)

Using more than two groups in an experiment also allows researchers to determine whether each treatment is more or less effective than no treatment (the control group). Suppose we are interested in the effects of aerobic exercise on anxiety. We hypothesize that the more aerobic activity one engages in, the more anxiety is reduced. We use a control group who does not engage in any aerobic activity and a high aerobic activity group who engages in 50 minutes per day of aerobic activity—a simple two-group design.

Now assume that when using this design, we find that both those in the control group and those in the experimental group have high levels of anxiety at the end of the study—not what we expected to find. A design with more than two groups might provide more information. Suppose we add another group to this study, a moderate aerobic activity group (25 minutes per day), and get the following results:

Control group	High anxiety
Moderate aerobic activity	Low anxiety
High aerobic activity	High anxiety

Based on these data, we have a V-shaped function. Up to a certain point, aerobic activity reduces anxiety. Yet when the aerobic activity exceeds a certain level, anxiety increases again. By conducting the study with only two groups, we miss this relationship and erroneously conclude that there is no relationship between aerobic activity and anxiety. Using a design with multiple groups allows us to see more of the relationship between the variables.

Figure 14.1 illustrates the difference between the results obtained with the three-group and with the two-group design in this hypothetical study. It also shows the other two-group comparisons: control compared to moderate aerobic activity and moderate aerobic activity compared to high aerobic activity. This set of graphs illustrates how two-group designs limit our ability to see the complete relationship between variables.

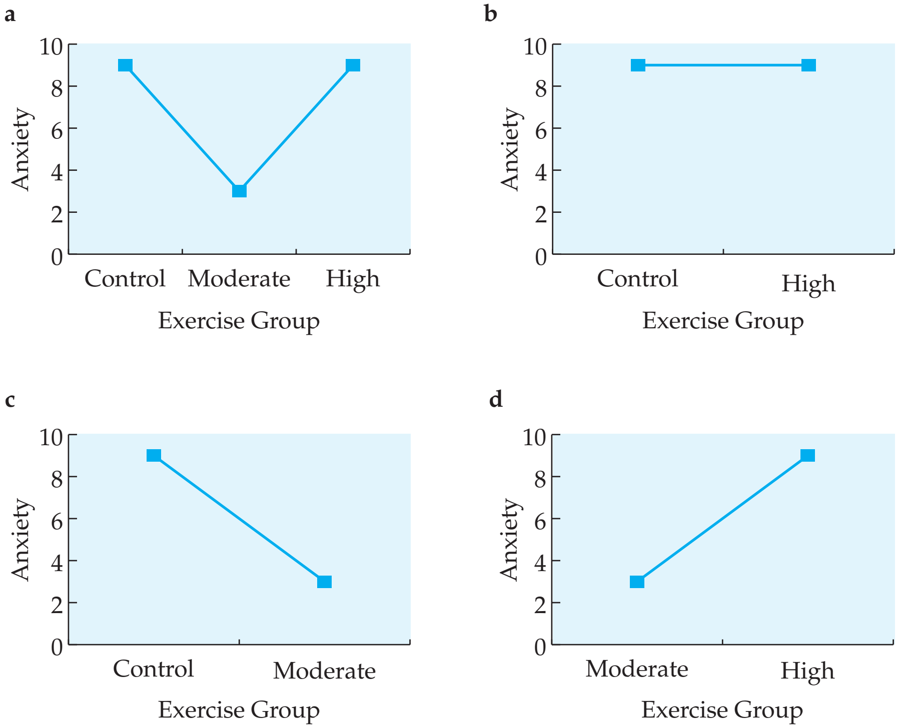

FIGURE **14.1** Determining relationships with three-group versus two-group designs: (a) three-groups design; (b) two-group comparison of control to high aerobic activity; (c) two-group comparison of control to moderate aerobic activity; (d) two-group comparison of moderate aerobic activity to high aerobic activity

Figure 14.1a shows clearly how the three-group design allows us to assess more fully the relationship between the variables. If we conduct only a two-group study such as those illustrated in Figure 14.1b, c, or d, we will draw a much different conclusion than that drawn from the three-group design. Comparing only the control to the high aerobic activity group (Figure 14.1b) leads us to conclude that aerobic activity does not affect anxiety. Comparing only the control and the moderate aerobic activity group (Figure 14.1c) leads us to believe that increasing aerobic activity reduces anxiety. Comparing only the moderate aerobic activity group and the high aerobic activity group (Figure 14.1d) lends itself to the conclusion that increasing aerobic activity increases anxiety.

Being able to assess the relationship between the variables means that we can determine the type of relationship. In the preceding example the variables produced a V-shaped function. Other variables may be related in a straight linear manner or in an alternative curvilinear manner (for example, a J- or

S-shaped function). In summary adding levels to the independent variable allows us to determine more accurately the type of relationship between variables.

Comparing a Placebo Group with the Control and Experimental Groups

A final advantage of designs with more than two groups is that they allow for the use of a *placebo group*, which improves an experiment. Consider an often-cited study by Paul (1966, 1967) involving children who suffered from maladaptive anxiety in public speaking situations. Paul used a control group, which received no treatment; a placebo group, which received a placebo that they were told was a potent tranquilizer; and an experimental group, which received desensitization therapy. Of those in the experimental group 85% showed improvement compared with only 22% in the control condition. If the placebo group had not been included, the difference between the therapy and control groups (85% − 22% = 63%) would have led to an overestimation of the effectiveness of the desensitization program. The placebo group showed 50% improvement, indicating that the therapy's true effectiveness is much less (85% − 50% = 35%). Thus a placebo group allows for a more accurate assessment of a therapy's effectiveness because in addition to spontaneous remission, it controls for participant expectation effects.

IN REVIEW Designs with More Than Two Levels of an Independent Variable

Advantages	Considerations
Allows comparisons of more than two types of treatment	Making multiple comparisons may lead us to draw an erroneous conclusion
Requires fewer participants	
Allows comparisons of all treatments with control condition	
Allows for use of a placebo group with control and experimental groups	

USING DESIGNS WITH MORE THAN ONE INDEPENDENT VARIABLE

We now turn to a discussion of more complex designs: those with more than one independent variable, or factor. As discussed above, these designs are usually referred to as factorial designs, indicating that more than one factor, or variable, is manipulated in the study. In the study of the effects of rehearsal on memory, participants used one of three types of rehearsal (rote, imagery, or story) to determine their effects on the number of words recalled. Imagine that, upon further analysis of the data, we discovered that participants recalled concrete words such as desk, bike, and tree better than abstract words such as love, truth, and honesty in one rehearsal condition but not in another. Such a result is called an *interaction* between variables (a concept discussed in more detail later in the module). One advantage of using factorial designs is that they allow us to assess how variables interact. In the real world it is unusual to find that a certain behavior is produced by only one variable; behavior is usually contingent on many variables operating together in an interactive way. Designing experiments with more than one independent variable allows researchers to assess how multiple variables may affect behavior.

Factorial Notation and Factorial Designs

A factorial design then is one with more than one factor, or independent variable. A *complete factorial design* is one in which all levels of each independent variable are paired with all levels of every other independent variable. An *incomplete factorial design* also has more than one independent variable, but all levels of each variable are not paired with all levels of every other variable. The design illustrated in this module is a complete factorial design.

An independent variable must have at least two levels because if it does not vary, it is not a variable. Consequently the simplest complete factorial design is one with two independent variables, each with two levels. Let's consider an example. Suppose we manipulate two independent variables: word type (concrete versus abstract words) and rehearsal type (rote versus imagery). The independent variable Word Type has two levels, abstract and concrete; the independent variable Rehearsal Type also has two levels, rote and imagery. This design is known as a 2×2 factorial design.

The factorial notation for a factorial design is determined as follows:

Number of levels of independent variable 1 $\times$ Number of levels of independent variable 2 $\times$ Number of levels of independent variable 3 and so on ...

factorial notation: The notation that indicates how many independent variables are used in a study and how many levels are used for each variable.

Thus the **factorial notation** indicates how many independent variables are used in the study and how many levels are used for each of them. This concept is often confusing for students, who frequently think that in the factorial notation 2×2 the first number (2) indicates that there are two independent variables and the second number (2) indicates that each has two levels. This is *not* how to interpret factorial notation. Rather each number in the notation specifies the number of levels of a single independent variable. So a 3×6 factorial design is one with two independent variables; each of the two numbers in the factorial notation represents a single independent variable. In a 3×6 factorial design one independent variable has three levels whereas the other has six levels.

Referring to our 2×2 factorial design, we see that there are two independent variables, each with two levels. This factorial design has four conditions ($2 \times 2 = 4$): abstract words with rote rehearsal, abstract words with imagery rehearsal, concrete words with rote rehearsal, and concrete words with imagery rehearsal. How many conditions would there be in a 3×6 factorial design? If you answer 18, you are correct. Is it possible to have a 1×3 factorial design? If you answer no, you are correct because it is not possible to have a factor (variable) with one level because then it does not vary.

Main Effects and Interaction Effects

main effect: An effect of a single independent variable.

interaction effect: The effect of each independent variable across the levels of the other independent variable.

Two kinds of information can be gleaned from a factorial design. The first is whether there are any main effects. A **main effect** is an effect of a single independent variable. In our design with two independent variables, two main effects are possible: an effect of word type and an effect of rehearsal type. In other words, there can be as many main effects as there are independent variables. The second type of information is whether there is an interaction effect, which as the name implies, is information regarding how the variables or factors interact. Specifically an **interaction effect** is the effect of each independent variable across the levels of the other independent variable. When there is an interaction between two independent variables, the effect of one independent variable depends on the level of the other independent variable. An example will make this point clearer.

Let's look at the data from the study on the effects of word type and rehearsal type on memory. Table 14.1 presents the mean performance for participants in each condition. This design was completely between-participants, with different participants serving in each of the four conditions. There were 8 participants in each condition, for a total of 32 participants in the study. Each participant in each condition was given a list of 10 words (either abstract or concrete) to learn using the specified rehearsal technique (rote or imagery).

Typically researchers begin by assessing whether there is an interaction effect because such an effect indicates that the effect of one independent variable depends on the level of the other. However, when first interpreting the results of two-way designs, students usually find it easier to look at the main effects and then move on to the interaction effect. Keep in mind that if we later find an interaction effect, then any main effects have to be qualified. Because we have two independent variables, there is the possibility of two main effects: one for word type (variable A in the table) and one for rehearsal type

TABLE **14.1**

Results of the 2 × 2 Factorial Design: Effects of Word Type and Rehearsal Type on Memory

	Word Type (Independent Variable A)		
Rehearsal Type (Independent Variable B)	**Concrete**	**Abstract**	**Row Means (Main Effect of B)**
Rote rehearsal	5	5	5
Imagery rehearsal	10	5	7.5
Column means (Main Effect of A)	7.5	5	

(variable B). The main effect of each independent variable tells us about the relationship between that single independent variable and the dependent variable. That is, do different levels of one independent variable bring about changes in the dependent variable?

We can find the answer to this question by looking at the row and column means in Table 14.1. The column means tell us about the overall effect of variable A (word type). They indicate that there is a difference in the numbers of words recalled between the concrete and abstract word conditions. More concrete words were recalled (7.5) than abstract words (5). The column means represent the average performance for the concrete and abstract word conditions summarized across the rehearsal conditions. In other words, we obtained the column mean of 7.5 for the concrete word conditions by averaging the numbers of words recalled in the concrete word/rote rehearsal condition and the concrete word/imagery rehearsal condition $[(5 + 10)/2 = 7.5]$. Similarly, the column mean for the abstract word conditions (5) was obtained by averaging the data from the two abstract word conditions $[(5 + 5)/2 = 5]$. (Note that determining the row and column means in this manner is possible only when the numbers of participants in each of the conditions are the same. If the numbers of participants in the conditions are unequal, then all individual scores in the single row or column must be used in the calculation of the row or column mean.)

The main effect for variable B (rehearsal type) can be assessed by looking at the row means. The row means indicate that there is a difference in the numbers of words recalled between the rote rehearsal and the imagery rehearsal conditions. More words were recalled when participants used the imagery rehearsal technique (7.5) than when they used the rote rehearsal technique (5). As with the column means the row means represent the average performance in the rote and imagery rehearsal conditions summarized across the word type conditions.

At face value the main effects tell us that overall participants recall more words when they are concrete and when imagery rehearsal is used. However, we now need to assess whether there is an interaction between the variables. If so, the main effects noted previously have to be qualified because an interaction indicates that the effect of one independent variable depends on the level of the other. That is, an interaction effect indicates that the effect of one independent variable varies at different levels of the other independent variable.

Look again at the data in Table 14.1. There appears to be an interaction in these results because when rote rehearsal is used, word type makes no difference (the means are the same—5 words recalled). Yet when imagery rehearsal is used, word type makes a big difference. Specifically then, when imagery is used with concrete words, participants do very well (recall an average of 10 words), yet when imagery is used with abstract words, participants perform the same as they did in both of the rote rehearsal conditions (they recall an average of only 5 words). Think about what this result means. When there is an interaction between the two variables, the effect of one independent variable differs at different levels of the other independent variable—there

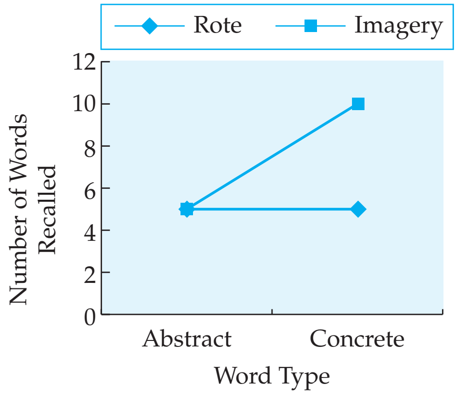

FIGURE **14.2** Line graph representing interaction between rehearsal type and word type

is a contrast or a difference in the way participants perform across the levels of the independent variables.

Another way to assess whether there is an interaction effect in a study is to graph the means. Figure 14.2 shows a line graph of the data presented in Table 14.1. The interaction may be easier for you to see here. First, when there is an interaction between variables, the lines are not parallel; they have different slopes. You can see in the figure that one line is flat (representing the data from the rote rehearsal conditions), whereas the other line has a positive slope (representing the data from the imagery rehearsal conditions). Look at the figure and think about the interaction. The flat line indicates that when rote rehearsal was used, word type had no effect; the line is flat because the means are the same. The line with the positive slope indicates that when imagery rehearsal was used, word type had a big effect; participants remembered more concrete words than abstract words.

Although the concept of interaction may seem difficult to understand, interactions often occur in our own lives. When we say "It depends," we are indicating that what we do in one situation depends on some other variable, that is, there is an interaction. For instance, whether you go to a party depends on whether you have to work and who is going to be at the party. If you have to work, you will not go to the party under any circumstance. However, if you do not have to work, you might go if a "certain person" is going to be there. If that person is not going to be there, you will not go. See if you can graph this interaction. The dependent variable, which always goes on the *y*-axis, is the likelihood of going to the party. One independent variable is placed on the *x*-axis (whether or not you have to work), and the levels of the other independent variable are captioned in the graph (whether the certain person is or is not present at the party).

To determine whether main effects or an interaction effect are meaningful, we need to conduct statistical analyses. We briefly discuss the appropriate analysis later in the text, but a more thorough discussion can be found in a statistics text.

Possible Outcomes of a 2 × 2 Factorial Design

A 2 × 2 factorial design has several possible outcomes. Because there are two independent variables, there may or may not be a significant effect of each. In addition, there may or may not be a significant interaction effect. Consequently there are eight possible outcomes in all (possible combinations of significant and nonsignificant effects). Figure 14.3, using the word recall study as an example, illustrates these eight possible outcomes for a 2 × 2 factorial design. Obviously only one of these outcomes is possible in a single study, but all eight are graphed to give a concrete illustration of each possibility.

For each graph the dependent variable (number of words recalled) is placed on the *y*-axis, and independent variable A (word type) is placed on the *x*-axis. The two means for one level of independent variable B (rehearsal type) are plotted, and a line is drawn to represent this level of independent variable B. In the same fashion the means for the second level of independent variable B are plotted, and a second line is drawn to represent this level of independent variable B. Next to each graph is a matrix showing the means from the four conditions in the study. The graphs are derived by plotting the four means from each matrix. In addition, whether there are main effects and an interaction effect is indicated.

Can you tell by looking at the graphs which ones represent interaction effects? Graphs a, b, c, and d do not have interaction effects, and graphs e, f, g, and h do. You should have a greater appreciation for interaction after looking at these graphs. Notice that in graphs a–d there is no interaction because each level of independent variable A (word type) affects the levels of independent variable B (rehearsal type) in the same way. Look at graphs c and d. In graph c the lines are parallel with no slope. This result indicates that for both rote and imagery rehearsal, word type makes no difference. In graph d the lines are parallel and sloped. This result indicates that for both rote and imagery rehearsal, word type has the same effect: Performance is poorer for abstract words and then increases by the same amount for concrete words.

Now look at graphs e–h, which represent interaction effects. Sometimes there is an interaction because even though there is no relationship between the independent variable and the dependent variable at one level of the second independent variable, there is a strong relationship at the other level of the second independent variable. Graphs e and f show this. In graph e when rote rehearsal is used, word type makes no difference, whereas when imagery rehearsal is used, word type makes a big difference. In graph f the interaction is due to a similar result. Sometimes, however, an interaction may indicate that an independent variable has an opposite effect on the dependent variable at different levels of the second independent variable. Graphs g and h illustrate this. In graph g when rote rehearsal is used, performance improves for concrete words versus abstract words (a positive relationship). However, when imagery rehearsal is used, performance decreases for concrete words versus abstract words (a negative relationship). Finally, graph h shows similar but more dramatic results. Here there is a complete crossover interaction in which exactly the opposite result is occurring for independent variable B at the levels of independent variable A. Notice also in this graph that although there is a large crossover interaction, there are no main effects.

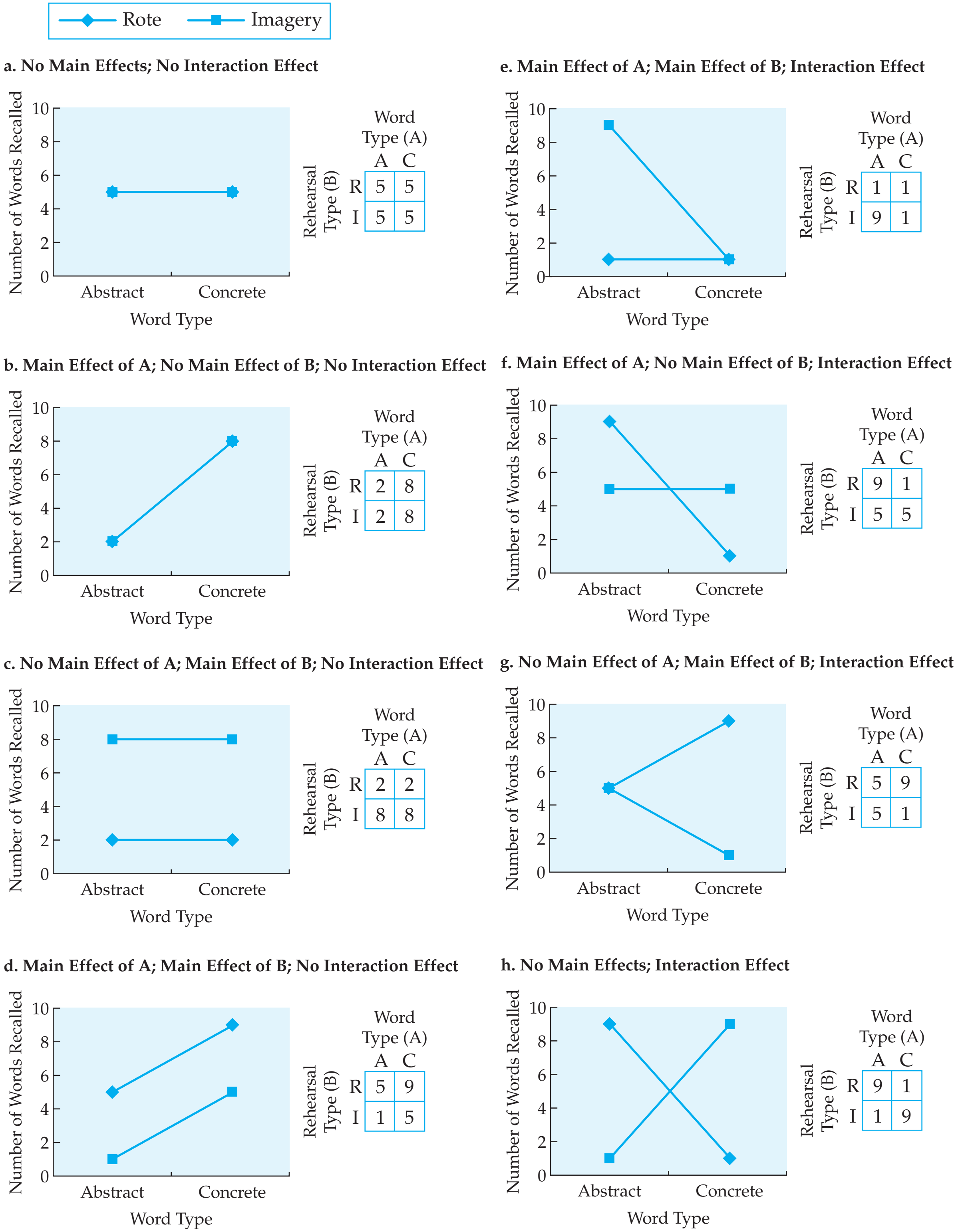

FIGURE **14.3** Possible outcomes of a 2 × 2 factorial design with rehearsal type and word type as independent variables

To make sure you completely understand interpreting main effects and interaction effects, cover the titles in each part of Figure 14.3 and quiz yourself on whether there are main effects and/or an interaction effect in each graph.

IN REVIEW Complex Designs

	Description	Advantage or Example
Factorial design	Any design with more than one independent variable	In the word recall study example, word type and rehearsal type were both manipulated to assess main effects and an interaction effect. The advantage is that the study more closely resembles the real world because the results are due to more than one factor (or variable).
Factorial notation	The numerical notation corresponding to a factorial design, indicating, in brief form, the number of independent variables and the number of levels of each variable	A 3 × 4 design has two independent variables, one with three levels and one with four levels.
Main effect	An effect of a single variable. A main effect describes the effect of a single variable as if there were no other variables in the study	In a study with two independent variables, two main effects are possible, one for each variable.
Interaction effect	The effect of each independent variable at the levels of the other independent variable	Interaction effects allow us to assess whether the effect of one variable depends on the level of the other variable. In this way we can more closely simulate the real world, where multiple variables often interact.

CRITICAL THINKING CHECK 14.1

1. What is the factorial notation for the following design? A pizza parlor owner is interested in which type of pizza his customers most prefer. He manipulates the type of crust for the pizzas by using thin, thick, and hand-tossed crusts. In addition, he manipulates the topping for the pizzas by offering cheese, pepperoni, sausage, veggie, and everything. He then has his customers sample the various pizzas and rate them. After you have determined the factorial notation, indicate how many conditions are in this study.
2. How many main effect(s) and interaction effect(s) are possible in a 4 × 6 factorial design?
3. Draw a graph representing the following data from a study using the same independent variables as in the module example.

(continues)