Chapter 13. Experimental Design: Multiple Independent Variables

Characteristics of Factorial Designs

Possible Outcomes of a 2 X 2 Factorial Experiment

Different Types of Factorial Designs

Completely randomized (independent samples)

Repeated measures

Mixed design

Interpreting Main Effects and Interactions

More Complex Factorial Designs

Case Analysis

General Summary

Detailed Summary

Key Terms

Review Questions/Exercises

Characteristics of Factorial Designs

Why do children engage in aggressive behaviors? From our discussions thus far, it is clear that aggression is the result of several factors. Indeed, nearly all behaviors have their cause in a multitude of factors, both genetic and environmental factors. Thus, when we attempt to understand some area of human or animal behavior, it is often advantageous to be able to study more than one independent variable at a time. In the previous two chapters, we discussed experimental designs that involved only one independent variable and one dependent variable. In this chapter, we examine factorial designs that include more than one independent variable.

In this book, we have continued to discuss the independent variable of TV violence and its potential impact on the aggressive behavior of children. One issue that we have not discussed, but others have, is whether the violence depicted on television programs is perceived as real. Does a child realize that violence seen in cartoons is not real? Might exposure to such violence affect the child's behavior in a way different from real characters? This is an interesting question. In the previous chapters, we would have treated the type of television character as an extraneous variable and used a design technique to control it (e.g., only used shows with real characters). Now, we are interested in treating the realism of the TV program (real characters versus cartoon characters) as a second independent variable and examining its possible effect on aggressive behavior.

As you will see, the primary advantage of this design is the ability to study more than one independent variable at a time and also how these variables interact with one another. For our study, we will examine (1) the effect of TV violence on aggressive behavior (violent vs nonviolent TV programs), (2) the effect of program realism on aggressive behavior (cartoon vs real characters), and (3) the possible interaction effect of TV violence and program realism on aggressive behavior. Notice that the first two effects could be tested by conducting two separate one-way designs. One study could test the effect of TV violence and the other could test the effect of program realism. However, two separate studies would not permit us to examine the possible interaction of these two variables. That is, could the effect of TV violence depend on whether the program characters are real? We will return to this experiment a little later in this chapter.

In a two-way factorial design, there are two experimental or treatment variables (independent variables). One or both of these variables may be either qualitative (distinct categories) or quantitative (different amounts). Although there are only two experimental variables in a two-way design, there may be any number of subclasses or levels of treatment of each variable. A given study might involve two levels of one variable and four levels of a second variable, or three levels of each variable, etc. The traditional way of designating a two-way design is by citing the number of levels (or subclasses) of each variable. Some examples are: (a) A study with two levels of one variable and four levels of a second variable is referred to as having a 2 X 4 design. This 2 X 4 study would have 8 cells (groups). (b) A study with three levels of each variable is referred to as having a 3 X 3 design. This design would have 9 groups. (c) A study with three levels of one variable and four levels of a second variable is referred to as having a 3 X 4 design (12 groups).

The only inherent limitation of designs with multiple independent variables is that it is a bit more complex than the one-way designs. Data analysis and interpretation are a little more challenging and the design often requires more participants than designs with one independent variable. This is particularly true as the design is expanded to include three or more independent variables or three or more levels for the independent variables. Fortunately for you, we will restrict our focus to two-way designs, i.e., only two independent variables.

Table 13.1 summarizes the experimental designs discussed thus far.

Table 13.1 Characteristics of Research Designs

Type of Research Design	Number of IVs	Number of Levels of the IV	Number of DVs	Assignment to Conditions	Most Probable Inferential Statistic
One-way independent samples	1	2 or more	1	Random	*t* test or one-way ANOVA
One-way correlated samples	1	2 or more	1	Natural pairs Matched pairs Repeated measures	*t* test or one-way ANOVA
Two way design	2 or more	2 or more	1	Random Repeated measures Mixed	Two-way ANOVA

Possible Outcomes of a 2 x 2 Factorial Experiment

The total number of treatment combinations in any factorial design is equal to the product of the treatment levels of all factors or variables. Thus, in a 2 X 2 factorial design, there are four treatment combinations and in a 2 X 3 factorial design there are six treatment combinations. In more complex factorial designs, the same principle applies. In a 2 X 3 X 4 factorial design, there are 24 treatment combinations.

As noted, factorial designs introduce the concept of interaction. The concept is very important for the proper analysis and understanding of complex designs. Indeed often the main interest of a study is focused on the interaction of variables, i.e., is the effect of one variable dependent on either the presence of the amount of a second variable? For example, a psychoactive drug alone may have little effect on the treatment of mentally ill patients. Psychotherapy alone may be equally ineffective. However, the combination of the two may produce the desired behavioral change. Without the two-variable design, this interaction might never be discovered.

Figure 13.1 illustrates a possible interaction of two variables – psychotherapy vs no psychotherapy and drug vs. no drug. Note that psychotherapy alone and the drug alone did not appear to bring about much improvement in the behavior of the patients. However, when both were combined, an improvement is noted (right side of top line). Let's look at another example. Imagine you had conducted a study involving two levels of anxiety (A_1 = low, A_2 = high) and two levels of perceived difficulty of task (B_1 = easy, B_2 = difficult). The dependent measure is the time the individual continues to work at an unsolvable task. Figure 13.2 illustrates six different, although not exhaustive, possible outcomes of the experiment. In (a) there is no effect of either independent variable and there is no interaction effect of the variables. In short, none of the experimental treatments had an effect on the dependent variable of time on task. Note that in (b), (c), and (d), the observed effect of one variable is found over both levels of a second variable. When this happens, we refer to the outcome as a main effect. A main effect of a given variable describes an effect of that variable over all levels of a second variable.

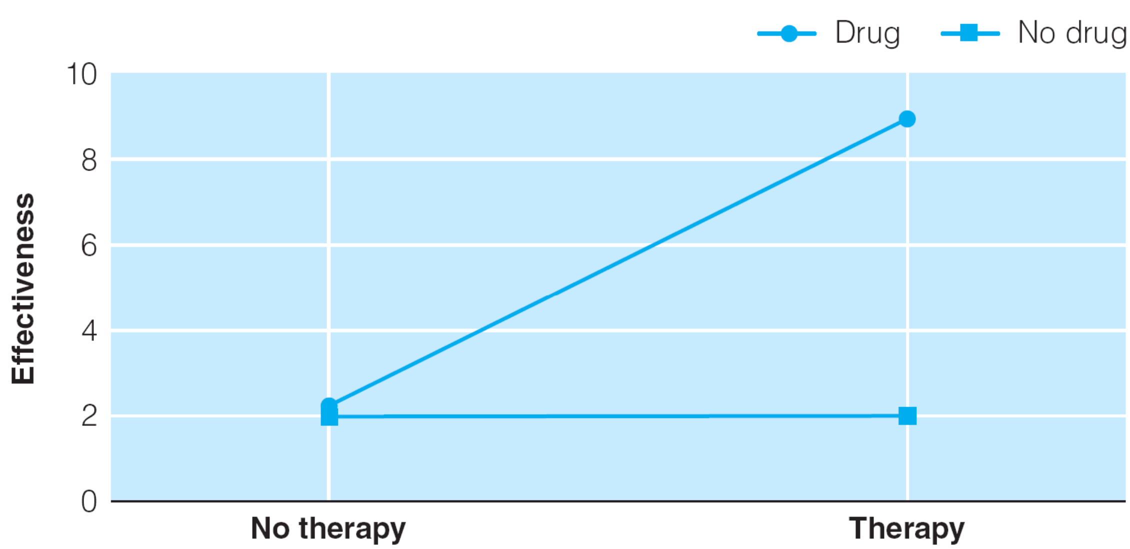

Figure 13.1 Interaction of psychotherapy treatment and drug treatment on therapeutic effectiveness

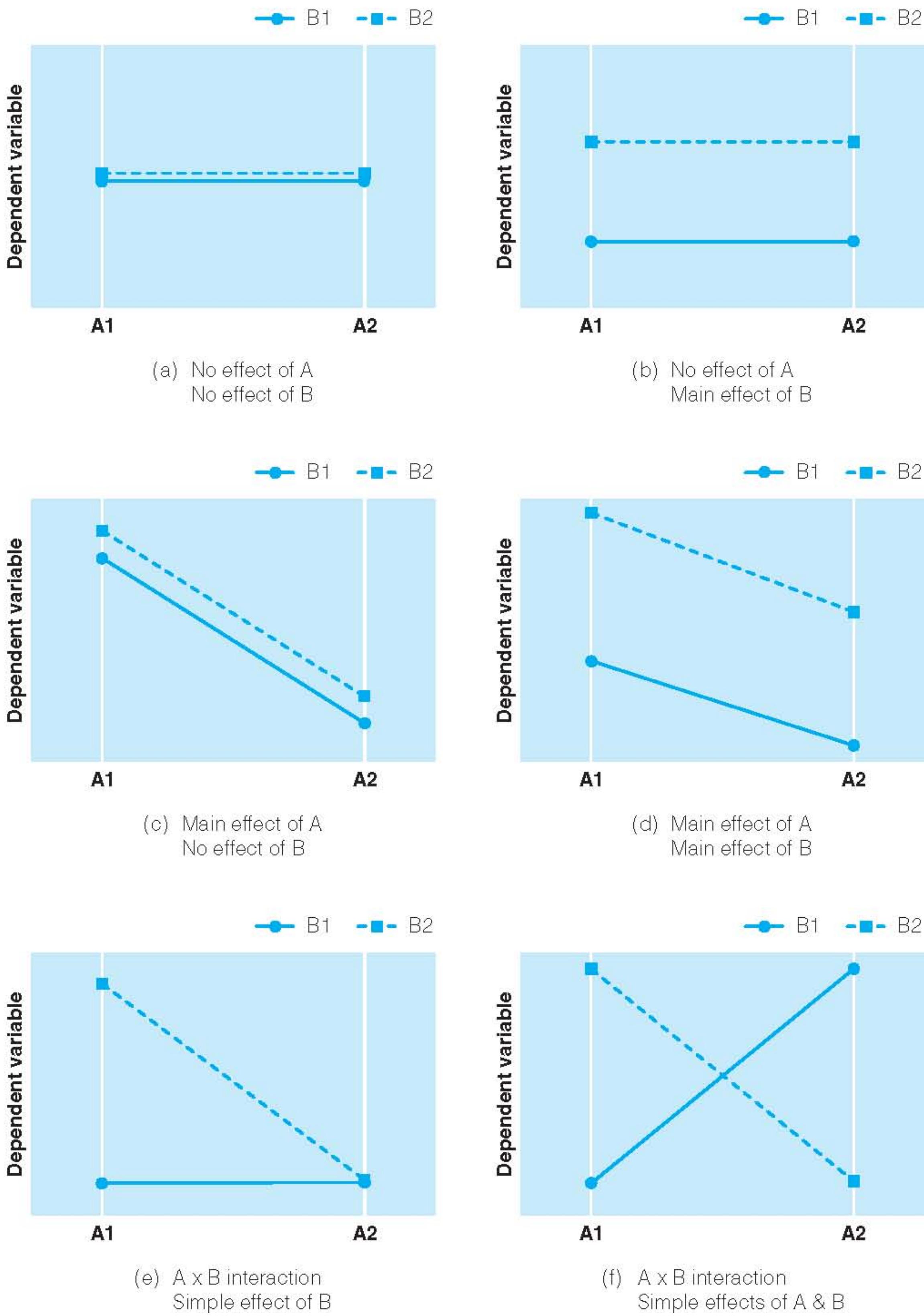

Figure 13.2 Several possible outcomes of an experiment involving a 2 X 2 factorial design.

Thus, in (b), high levels of perceived difficulty (B_2) is found to produce longer periods of task-related activities at both levels of induced anxiety. Similarly, in (c), the effect of low levels of anxiety is greater than high levels whether or not the task is perceived as easy (B_1) or difficult (B_2). Both of these are main effects. In contrast, (e) and (f) show interactions between the two. In (e), a task perceived as difficult (B_2) increases the time spent at the task only when induced anxiety is low (A_1). In (f), the interaction is complete. A high level of perceived difficulty produces a greater effect than a low level only when induced anxiety is high (A_2). Here, a low level of perceived difficulty produces longer periods of task orientation only when induced anxiety (A_2) is high. Because the difference between low and high perceived difficulty is not found at both levels of induced anxiety, the effect is a simple effect.

Different Types of Factorial Designs

The previous two chapters discussed several different ways that participants can be assigned to experimental conditions. These methods included randomization, natural pairs, matched pairs, and repeated measures. These options continue to be available to us in the two-way design.

Completely randomized factorial design (independent samples)

A completely randomized factorial design uses randomization to assign participants to all treatment conditions. Let's consider the use of a 2 X 2 factorial design for our TV violence study. Participants will be randomly assigned to one of the levels of TV violence and one of the levels of program realism. Thus, in a 2 X 2 factorial design, there are four independent groups and participants are randomly assigned to one of the four groups. Table 13.2 shows both the general terms associated with this design and the specific labels for our study.

Table 13.2 General Diagram and Specific Example of a 2 × 2 Factorial Design

		FACTOR A (TV VIOLENCE)	
		Level 1 (A1) (Violent)	**Level 2 (A2) (Nonviolent)**
FACTOR B (REALISM)	**Level 1 (B1) (Real)**	A1B1 (Real/Violent)	A2B1 (Real/Nonviolent)
	Level 2 (B2) (Cartoon)	A1B2 (Cartoon/Violent)	A2B2 (Cartoon/Nonviolent)

Repeated measures design

A repeated measures design uses multiple observations on the same participants, possibly in combination with natural pairs or matched pairs, to assign participants to treatment conditions. The most typical design involves all participants participating in all conditions. Let's look at how a two-way repeated measures design might be used in the TV violence study.

Like the factorial design described earlier, we will have two independent variables, each with two levels. But rather than randomly assigning participants to each of the four conditions, we will now have all participants participate in all four conditions. Table 13.3 illustrates this.

Table 13.3 Example of a Repeated Measures Factorial Design

	Experimental Condition			
	Real/ Violent	Real/ Nonviolent	Cartoon/ Violent	Cartoon/ Nonviolent
Participant 1				
Participant 2				
Participant 3				
.				
.				
.				
Participant 12				

As discussed previously, advantages of the repeated measures design include reduced error due to individual differences and the efficiency obtained by requiring fewer participants. Disadvantages include concern for homogeneity of variance, homogeneity of covariance, and carryover effects. Regarding carryover effects, it may be that the effect of watching a violent program may carryover into an observation period following a nonviolent program. Although the statistical concerns are difficult to control, carryover effects can be addressed by counterbalancing the conditions.

For our example with four different conditions, complete counterbalancing would require 24 different orders for the conditions. In many situations, the experimenter is unable to accommodate this. One possible solution is to select a random order for each participant. Another technique uses a simpler version in which each condition occurs in a different position in the order. For example, if we label the real/violent condition as condition 1, the real/nonviolent as condition 2, etc., then we would arrange the ordering as shown in Table 13.4.

Table 13.4 Counterbalancing in a Repeated Measures Factorial Design

	Experimental Condition			
Participant 1	1	2	3	4
Participant 2	1	2	3	4
Participant 3	1	2	3	4
Participant 4	2	3	4	1
Participant 5	2	3	4	1
Participant 6	2	3	4	1
Participant 7	3	4	1	2
Participant 8	3	4	1	2
Participant 9	3	4	1	2
Participant 10	4	1	2	3
Participant 11	4	1	2	3
Participant 12	4	1	2	3

Thus, each condition occurs in the first position three times, each condition occurs in the second position three times, etc. One potential danger of this simpler counterbalancing occurs if the effect of one condition depends on the particular condition that preceded it in the order.

Mixed design

A mixed design uses a combination of randomization and repeated measures (although natural pairs and matched pairs are possible) to assign participants to treatment conditions. Participants are randomly assigned to the different levels of one independent variable and participate in all levels of another independent variable. For our TV violence study, we might decide to randomly assign participants to watch either TV programs with violence or ones without violence, but all participants will watch one show involving real characters and one show involving cartoon characters. We might visualize this design as shown in Table 13.5.

Table 13.5 Example of a Mixed Factorial Design

			REALISM	
			Real	Cartoon
TV VIOLENCE	Violent	Participant 1		
		Participant 2		
		Participant 3		
		Participant 4		
		Participant 5		
		Participant 6		
		Participant 7		
		Participant 8		
	Nonviolent	Participant 9		
		Participant 10		
		Participant 11		
		Participant 12		
		Participant 13		
		Participant 14		
		Participant 15		
		Participant 16		

Let's turn to a recent report of the effects of diet on blood pressure and consider how the study might be converted into a factorial design (see Box 13.1).

Box 13.1 Thinking Critically About Everyday Information – Effect of diet on blood pressure

A report by ABC News refers to research (Appel et al., 2003) that studied the effect of lifestyle changes on lowering blood pressure. In the study, participants with above-normal blood pressure were randomly assigned to one of three treatment conditions. One group received one advice session regarding dietary changes, the second group received 28 counseling sessions that focused on lowering fat intake, and the third group received the 28 counseling sessions plus a specific diet plan. After six months, results showed that substantial improvements in both the second and third groups.

1. What type of research design was used?
2. Was the study a true experiment? Why or why not?
3. As you know, lifestyle change can involve more than changes in diet. What second independent variable could be added to create a factorial design? What levels of this second IV would you test?
4. What type of factorial design do you now have (completely randomized, repeated measures, or mixed design)?
5. What would be your prediction for your new factorial design?

Retrieved June 11, 2003 online at: http://abcnews.go.com/sections/living/Healthology/ho_bpsqueeze.html

Interpreting Main Effects and Interactions

Let's return to our two-way factorial design to discuss the interpretation of main effects and interactions. Although we won't do the same for two-way repeated measures or mixed designs, the logic of the interpretation is the same.

Assume that we began with a sample of 80 participants, randomly assigned 20 to each of the four groups, and calculated the mean number of aggressive behaviors for each group. Table 13.6 shows the design and descriptive statistics.

Table 13.6 Example of a 2 × 2 Factorial Design

		TV Violence	
		Violent	Nonviolent
Characters	Real	M = 7.75	M = 3.95
		SD = 2.63	SD = 1.32
	Cartoon	M = 5.95	M = 4.00
		SD = 1.88	SD = 1.38

To aid interpretation, Figure 13.3 provides both a bar graph and a line graph of the means. Based on the graphs, can you guess which effects might be significant?

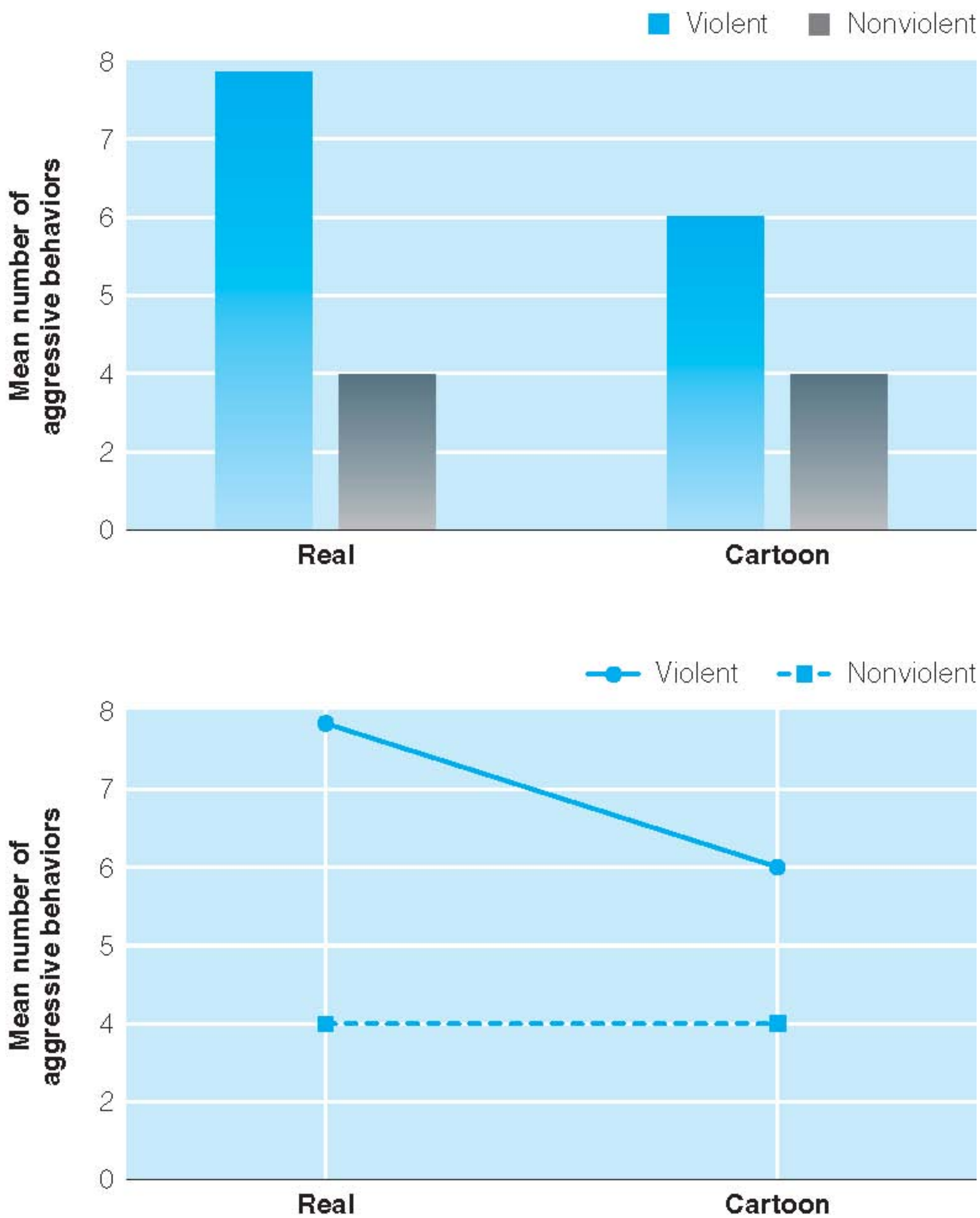

Figure 13.3 Graphical depictions of the means for the TV violence study.

Inferential analysis consisted of a 2 X 2 independent samples ANOVA. The ANOVA produced ANOVA Summary Table shown in Table 13.7.

Table 13.7 ANOVA Summary Table

Source of Variability	Degrees of Freedom (*df*)	Sum of Squares	Mean Squares	*F* Ratio	*F* Probability
Level of Violence	1	165.31	165.31	46.94	<.001
Type of Characters	1	15.31	15.31	4.35	.040
Violence * Characters	1	17.11	17.11	4.86	.031
Error	76	3.52	3.52		
Total	79	201.25			

Notice that there are three F-values in the analysis – one for each effect to be analyzed. The analysis (assuming alpha = .05) reveals a significant main effect for the level of violence, a significant effect for the type of characters, and a significant interaction, $F(1,76) = 46.94$, $p < .001$, $F(1,76) = 4.35$, $p = .04$, and $F(1,76) = 4.86$, $p = .031$ respectively. Let's work to understand each of these effects individually. Using the graphs in Figure 13.3, the main effect for the level of violence shows that a program with violence resulted in significantly more aggressive behaviors than a program without violence. Notice that this interpretation ignores the levels of the other variable (type of characters). The main effect for the type of characters shows that real characters resulted in significantly more aggressive behaviors than cartoon characters. Notice that this interpretation ignores the levels of the other variable (level of violence). The interaction effect shows that the increase in aggressive behavior as a result of TV violence was enhanced when the TV characters were real. Notice that this interpretation considers both levels of both variables.

Are we done with the interpretation of results? Not yet. Whenever a significant interaction is found, you must determine whether the conclusions for the main effects still make sense. In other words, a statistically significant main effect can be explained by a significant interaction. So, let's return to each of our main effects. We concluded that TV violence resulted in more aggression. Does examination of our graphs show this to be always true? The answer is yes. TV violence resulted in more aggression when the TV program contained either type of character. We also concluded that real characters resulted in more aggression. Does examination of our graphs show this to be always true? The answer is no. Real characters only resulted in more aggression when the TV program contained violence. Thus, this main effect is actually a simple effect and a result of the significant interaction. Our overall conclusion would be that watching a TV program with violence resulted in more aggressive behaviors and this effect was enhanced by the use of real characters in the TV program.

It is important to notice in the above example, as well as the other examples in this chapter, that the researcher was able to randomly assign participants to the various conditions and was able to manipulate the factors of interest (independent variables). As we have noted before, these are important characteristics of a true experiment and permit cause/effect conclusions. As we stated in Chapter 11, students are often interested in factors such a gender, race, personality traits, political affiliation, fraternity/sorority status, etc where random assignment is not possible – designs in which one or more factors consists of preexisting groups. Again, you should exercise caution when drawing conclusions regarding the effects that result when being unable to randomly assign participants to conditions.

More Complex Factorial Designs

Before we end this chapter, we would like to say a few words about factorial designs that go beyond the relatively simple 2 X 2 design. The logic of main effects and interactions can be extended to 2 X 3 designs, 2 X 4 designs, 3 X 5 designs, etc. and these designs can be of the three different varieties presented in this chapter, namely independent samples, correlated samples, or mixed designs. With these more complex two-way designs, significant effects must often be followed by more specific comparison tests to determine exactly where significant differences exist. As illustrated earlier in this chapter, graphic depiction of the cell means is critical to a complete understanding of the data.

In addition to two-way designs, the logic of factorial designs can be extended to three-way designs (e.g., 2 X 2 X 2,; 2 X 4 X 3). Although the logic could be extended to four-way designs, five-way designs, and beyond, very few researchers employ such complicated experimental designs. In a three-way design, the ANOVA would analyze the main effect of factor A, the main effect of factor B, the main effect of factor C, the A X B interaction, the A X C interaction, the B X C interaction, and the A X B X C interaction. If the complexity of such an analysis seems daunting, we certainly understand.

Case Analysis

One fascinating area of psychological research involves understanding the brain mechanisms responsible for learning and memory. Both human and animal research suggests that a structure in the brain called the hippocampus is important for the formation of certain types of memory, including spatial memory (memory for where things are located in your environment). In one research paradigm, rats are placed in a pool of milky water. At a particular location in this pool, there is a platform that is hidden just beneath the surface of the milky water. When the rat locates this platform, it will climb onto it so that it can get most of its body out of the water (as you might guess, rats don’t like water). After several trials in the pool, a normal rat will locate the platform quickly by using the spatial cues present in the room (unique items hanging on the wall, etc.).

As a researcher, you are interested in the effect of hippocampal damage on spatial learning and also whether the hippocampus is necessary for the retrieval of spatial memories that have already been learned.

To do this, you randomly assign rats to one of four groups. Group 1 are rats with no damage to the hippocampus and no prior experience in the pool. Group 2 are rats with no damage to the hippocampus and prior experience in the pool. Group 3 are rats with damage to the hippocampus and no prior experience in the pool. Group 4 are rats with damage to the hippocampus and prior experience in the pool (experience was prior to damage). For each rat, you measure the time it takes for the rat to locate the platform. The group means are:

Group 1	30 seconds
Group 2	5 seconds
Group 3	29 seconds
Group 4	32 seconds

Critical Thinking Questions

1. What are the independent variables and the dependent variable for this experiment?
2. What type of research design is this?
3. Create a 2 x 2 table that depicts the conditions and means.
4. Create a line graph that depicts the conditions and means.
5. Based on the table and graph, does there appear to be:
   - a significant main effect of “damage”?
   - a significant main effect of “experience”?
   - a significant interaction?

   (Note: your answers to #5 are guesses. You would need to know the variability in the groups and the results of the ANOVA to make definitive judgments.)
6. Based on your answers to #5 and your evaluation of simple effects, write a conclusion for the experiment.

General Summary

Factorial designs permit the researcher to determine the effect of more than one independent variable on a dependent variable and to determine the possible interaction of multiple independent variables. That is, the effect one independent variable may differ across different levels of another independent variable. With an independent samples design, participants are randomly assigned to levels of each independent variable. With a correlated samples design, participants are usually repeatedly measured in that they participate in each level of each independent variable. With a mixed designed, participants are randomly assigned to the levels of one independent variable and are repeatedly measured on the levels of the other independent variable.

A 2 X 2 factorial design requires interpretation of three effects – a main effect for the first IV, a main effect for the second IV, and the interaction of the two IVs. A main effect signifies that there is an effect of that IV on the DV regardless of the levels of the other IV. An interaction signifies that the effect of one IV depends on the levels of the other IV. In some cases, a significant main effect may be solely due to the interaction. We refer to this as a simple effect and interpretation of this effect must be done with caution. In the next chapter we will explore how to design an experiment that involves only a single participant and how we can systematically manipulate an independent variable across time to determine its effect on a dependent variable.

Detailed Summary

1. Factorial designs permit the systematic investigation of two or more independent variables in a single study.
2. According to conventional terminology, a 2 X 3 factorial design would have two levels of one IV and 3 levels of the other IV.
3. Disadvantages of factorial designs can include the complexity of design and interpretation and, at times, the increased number of participants necessary to conduct the study.
4. A factorial design will reveal whether there is an effect of each independent variable on the dependent variable (main effects) plus whether there is an interaction effect. A main effect indicates that one IV has an effect on the DV regardless of the levels of a second IV. An interaction indicates that the effect of one IV depends on the levels of a second IV.
5. Different types of factorial designs can be distinguished based on how participants are assigned to treatment conditions. The three most common types are the completely randomized design, the repeated measures design, and the mixed design.
6. A completely randomized factorial design involves the random assignment of participants to all treatment conditions (i.e., all levels of all IVs). A repeated measures design involves testing each participant in each and every treatment condition. A mixed design involves a combination of random assignment to the levels of one IV and repeated testing across the levels of another IV.
7. A factorial design is analyzed with a factorial ANOVA that will calculate an F-value for each main effect and for the interaction. The statistical significance of these effects, along with a close inspection of a graph of the means, provides an accurate conclusion for the study.
8. The logic of the 2 X 2 factorial design can be extended to designs in which there are more than two levels of an independent variable (e.g., 2 X 3, 3 X 4) and/or more than two independent variables (e.g., 2 X 2 X 2; 2 X 3 X 4).

Key Terms

Completely randomized (factorial) design

Interaction

Main effect

Mixed design

Repeated measures design

Simple effect

Review Questions/Exercises

1. Briefly describe a memory recall study that would be a 2 X 2 independent samples design in which the independent variables are mode of sensory experience (see the object vs. touch the object) and type of object (natural vs. man-made).
2. Briefly describe a memory recall study that would be a 2 X 2 correlated samples (repeated measures) design in which the independent variables are mode of sensory experience (see the object vs. touch the object) and type of object (natural vs. man-made).
3. Briefly describe a memory recall study that would be a 2 X 2 mixed design in which the independent variables are mode of sensory experience (see the object vs. touch the object) and type of object (natural vs. man-made).

For questions 4-8, consider the following experiment:

An experimenter wanted to test the hypothesis that males are more creative than females. She also hypothesized that the male superiority in creativity would be heightened under conditions involving ego. She manipulated ego involvement by telling half of the males and females that the task was a measure of intelligence and that their scores would be posted on a bulletin board (high ego involvement). She told the other half of the males and females that she wanted to test the reliability of a task she was developing and that they shouldn't put their names on the answer sheets (low ego involvement). Her test of creativity was an "unusual uses" test in which a person is given the name of an object (army compass, monkey wrench) and has to write as many different unusual uses for that object as he or she can in 5 minutes. Twenty-five males and 25 females were randomly assigned to each of the two ego-involvement conditions. The males were members of a senior ROTC class, and the females came from sorority pledge classes. All participants were given 5 minutes to write down as many unusual uses as they could for each object. Results showed that mean number of unusual uses written by males was 6.1 under low ego involvement and 9.2 under high ego involvement. Mean number of unusual uses written by females was 2.3 under low ego involvement and 2.3 under high ego involvement.

4. How many independent variables were used? How many dependent variables?
5. How would you describe the experimental design that was used?
6. Does the pattern of means suggest a main effect of Sex? A main effect of Ego Involvement? A Sex X Ego Involvement interaction?
7. What is the primary threat to internal validity?
8. What is the primary threat to external validity?