CHAPTER 8
Experimental Design II: Factorial Designs

Preview & Chapter Objectives

Chapter 7 introduced you to some basic experimental designs—those involving a single independent variable, with either two or more levels being compared. The next logical step is to increase the number of independent variables. When that happens, the result is a factorial design, the focus of this chapter. When you complete this chapter, you should be able to:

• Describe factorial designs using a standardized notation system (2 × 2, 3 × 5, etc.).
• Place data accurately into a factorial matrix.
• Understand what is meant by a main effect and know how to determine if one exists.
• Understand what is meant by an interaction effect and know how to determine if one exists.
• Identify the varieties of factorials that correspond to the single-factor designs of Chapter 7 (independent groups, matched groups, nonequivalent groups, repeated measures).
• Identify a mixed design and understand why counterbalancing is not always necessary in such a design.
• Identify a P × E factorial, recognize its historical connection to Kurt Lewin and understand what it means when such a design produces main effects and interactions.
• Calculate the number of participants needed to complete each of the factorial varieties.
• Know how to be an ethically competent experimenter.

As you have been working your way through this research methods course, you have probably noticed that experimental psychologists sometimes seem to have a language of their own. They talk about operationalizing constructs, building stem and leaf displays, and eliminating confounds, and when they talk about regression they are not discussing Freud. You haven’t seen anything yet. After this chapter, you’ll be able to say things like this: “It was a two by three mixed factorial that produced one main effect for the repeated measures variable plus an interaction.” Let’s start with the basics.

Factorial Essentials

Suppose you are interested in memory and wish to find out if recall can be improved by training people to use visual imagery while memorizing a list of words. You could create a simple two-group experiment in which some people are trained to use visual imagery techniques while memorizing the words and some are told to use rote repetition. Suppose you also wonder about how memory is affected by a word list’s presentation rate. Again, you could do a simple two-group study. Some participants see the lists at the rate of 2 seconds per word, others at 4 seconds per word. With a factorial design, both of these studies can be done as part of the same experiment.

By definition, a factorial design involves any study with more than one independent variable (also referred to as a “factor”). In principle this could involve dozens of variables, but in practice these designs usually involve two or three factors, or sometimes four. Let’s stay with the memory example as a way of introducing a notation system for describing factorials.

Identifying Factorial Designs

First, a factorial is described with a numbering system that simultaneously identifies the number of independent variables and the number of levels of each variable. Thus a 2 × 3 (read this as “two by three”) factorial design has two independent variables; the first has two levels, the second has three. A 3 × 4 × 5 factorial has three independent variables, with three, four, and five levels, respectively. The memory
study would be a $2 \times 2$ design, with two levels of the “type of training” independent variable (imagery and rote repetition) and two levels of the “presentation rate” independent variable (2 and 4 seconds per item).

Second, the conditions to be tested in a factorial study can be identified by looking at all possible combinations of the different levels of each independent variable. In the memory study, this produces a display called a factorial matrix, which looks like this:

<table>
<thead>
<tr>
<th>Type of training</th>
<th>Presentation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imagery</td>
<td>2-sec/word</td>
</tr>
<tr>
<td>Imagery/2 sec</td>
<td>Imagery/4 sec</td>
</tr>
<tr>
<td>Rote</td>
<td>4-sec/word</td>
</tr>
<tr>
<td>Rote/2 sec</td>
<td>Rote/4 sec</td>
</tr>
</tbody>
</table>

Before going on, there’s something you should note very carefully. Up to this point in the book, I have been using the concepts “conditions of the experiment” and “levels of the independent variable” as if they meant the same thing. These concepts indeed are interchangeable in single-factor experiments. In factorials, however, this is no longer the case. In all experimental designs, the term “levels” refers to the number of levels of the independent variable. In factorial designs, the term “conditions” equals the number of cells in the matrix like the one you just examined. Hence, the $2 \times 2$ memory study has two independent variables, each with two levels. It has four different conditions, however, one for each of the four cells. The number of conditions in any factorial can be determined simply by calculating the product of the numbers in the notation system. A $3 \times 3$ design has 9 conditions; a $2 \times 2 \times 4$ design has 16.

You can visualize a generalized $2 \times 2$ factorial matrix this way:

<table>
<thead>
<tr>
<th>Factor B</th>
<th>Level B1</th>
<th>Level B2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>condition</td>
<td>condition</td>
</tr>
<tr>
<td>Factor A</td>
<td>A1B1</td>
<td>A1B2</td>
</tr>
<tr>
<td></td>
<td>condition</td>
<td>condition</td>
</tr>
<tr>
<td>Level A1</td>
<td>A2B1</td>
<td>A2B2</td>
</tr>
</tbody>
</table>

It’s important to be clear about this labeling system because when you are using a computerized statistics package, this is the language you will probably encounter. If the computer asks you for the data from cell A2B1 and you enter the data from cell A1B2 by mistake, the analysis will proceed and give you a nice printout, but the results will be wrong. Obviously, it is essential to enter the data in the proper cells. You will encounter this labeling system if you work your way through the 2-way ANOVA in Appendix C.

Table 8.1 shows you how this system of laying out factorials and labeling the cells works with a $2 \times 4$ and a $2 \times 2 \times 2$ design. Ignore the matrices with shaded cells for the moment; those will make sense after you finish reading the next section.
**Table 8.1** Sample Factorial Designs

1. **2 × 4** factorial:

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
</tr>
</thead>
</table>
   A1 | A1B1 | A1B2 | A1B3 | A1B4 |
   A2 | A2B1 | A2B2 | A2B3 | A2B4 |

   Testing for the main effect of A (i.e., comparing A1 and A2)

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
</tr>
</thead>
</table>
   A1 | A1B1 | A1B2 | A1B3 | A1B4 |
   A2 | A2B1 | A2B2 | A2B3 | A2B4 |

   Testing for the main effect of B (i.e., comparing B1, B2, B3, and B4)

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
</tr>
</thead>
</table>
   A1 | A1B1 | A1B2 | A1B3 | A1B4 |
   A2 | A2B1 | A2B2 | A2B3 | A2B4 |

2. **2 × 2 × 2** factorial:

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
</table>
   A1 | A1B1C1 | A1B2C1 | A1B1C2 | A1B2C2 |

   Testing for the main effect of A (i.e., comparing A1 and A2)

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
</table>
   A1 | A1B1C1 | A1B2C1 | A1B1C2 | A1B2C2 |

   Testing for the main effect of B (i.e., comparing B1 and B2)

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
</table>
   A1 | A1B1C1 | A1B2C1 | A1B1C2 | A1B2C2 |

   Testing for the main effect of C (i.e., comparing C1 and C2)

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
</table>
   A1 | A1B1C1 | A1B2C1 | A1B1C2 | A1B2C2 |
In factorial studies, two kinds of results occur—main effects and interactions. Main effects refer to the overall influence of the independent variables, while interactions examine whether the variables combine to form a more complex result. Let's examine each in more detail.

Main Effects

In the memory experiment we've been using as a model, the researcher is interested in the effects of two independent variables: type of training and presentation rate. In factorial designs, the term main effect is used to describe the overall effect of a particular independent variable. So in a study with two independent variables, such as a $2 \times 2$ factorial, there can be at most two main effects. Determining the main effect of one factor involves using the data for all levels of the other factor(s). In the memory study, this can be illustrated as follows. The main effect of type of training is determined by combining the data for both presentation rates. Hence, all of the information in the lightly shaded cells (imagery) would be combined and compared with the combined data in the heavily shaded cells (rote):

<table>
<thead>
<tr>
<th>Type of training (A)</th>
<th>Presentation rate (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-sec/word</td>
</tr>
<tr>
<td>Imagery A1</td>
<td>imagery/2 sec A1B1</td>
</tr>
<tr>
<td>Rote A2</td>
<td>rote/2 sec A2B1</td>
</tr>
</tbody>
</table>

Similarly, the main effect of presentation rate is determined by combining the data for both types of training. In the following matrix, the effect of presentation rate would be evaluated by comparing all of the information in the lightly shaded cells (2 sec/item) with all the data in the heavily shaded cells (4 sec/item):

<table>
<thead>
<tr>
<th>Type of training (A)</th>
<th>Presentation rate (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-sec/word</td>
</tr>
<tr>
<td>Imagery A1</td>
<td>imagery/2 sec A1B1</td>
</tr>
<tr>
<td>Rote A2</td>
<td>rote/2 sec A2B1</td>
</tr>
</tbody>
</table>
If you now take a second look at the shaded matrices in Table 8.1, you'll see that I have indicated which cells combine during the various main effects analyses for the $2 \times 4$ and the $2 \times 2 \times 2$ designs.

Let's consider some hypothetical data that might be collected in a memory experiment like the example we've been using. Assume there are 25 participants in each condition (cell) and their task is to memorize a list of 30 words. The average number of words recalled for each of the four conditions might look like this:

<table>
<thead>
<tr>
<th>Type of training</th>
<th>Presentation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-sec/word</td>
</tr>
<tr>
<td>Imagery</td>
<td>17</td>
</tr>
<tr>
<td>Rote</td>
<td>12</td>
</tr>
</tbody>
</table>

Does imagery training produce better recall than rote repetition? That is, is there a main effect of type of training? The way to find out is to compare all the “imagery” data with all the “rote” data. Specifically, this involves calculating what are called “row means.” The “imagery” row mean is $20$ words $\frac{(17+23)}{2} = \frac{40}{2} = 20$, and the “rote” row mean is $15$ words $\frac{(12+18)}{2} = \frac{30}{2} = 15$. When asking if there is a main effect of training type, the question is: “Is the difference between the means of 20 and 15 statistically significant or due to chance?”

In the same fashion, calculating column means allows us to see if there is a main effect of presentation rate. For the 2 sec/item column, the mean is $14.5$ words; it is $20.5$ words for the 4 sec/item row (you should check this). Putting all of this together yields this outcome:

<table>
<thead>
<tr>
<th>Presentation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Imagery</td>
</tr>
<tr>
<td>Rote</td>
</tr>
<tr>
<td>Overall</td>
</tr>
</tbody>
</table>

For these data, then, it appears that imagery improves memory ($20 > 15$) and that recall is higher if the words are presented at a slower rate ($20.5 > 14.5$). That is, there seem to be two main effects here (of course, it takes an ANOVA to make a judg-
ment about whether the differences are significant or due to chance. For a real example of a study that produced a strong main effect, consider this case study of how visual imagery can enhance memory.

Case Study 14—A Significant Main Effect
The hypothetical example that I’ve been using, in which visual imagery is used to improve memory, wasn’t just pulled out of thin air. Researchers in cognitive psychology have known for some time that recall performance can be improved by using imagery when memorizing. A study by Wollen, Weber, and Lowry (1972) tried to evaluate separately two factors that might contribute to imagery’s usefulness. Some prior research had suggested that memory is better if the memorizer uses bizarre rather than normal images, presumably on the grounds that something really unusual will stand out in one’s mind. Other research demonstrated that memory could be enhanced if images for items to be remembered could be combined into a single image. Wollen et al. used a factorial design to examine both the bizarreness factor and the combination factor. Participants were given word pairs to study. For example, one pair was “piano-cigar.” During recall, their task was to respond “cigar” when given the word “piano.”

During presentation of the word pairs, participants were shown one of four different sets of drawings, including the ones in Figure 8.1, for the piano-cigar pair. As you can see, two of the images are combined and two are not, and two of the images are unusual (i.e., bizarre) and two are not. The mean number of pairs recalled in the four conditions of this 2 x 2 independent groups factorial showed a strong main effect for the combination factor, but no significant effect for the bizarreness factor (maximum score = 9):

<table>
<thead>
<tr>
<th></th>
<th>Bizarre</th>
<th>Normal</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>6.67</td>
<td>6.60</td>
<td>6.64</td>
</tr>
<tr>
<td>Not combined</td>
<td>3.05</td>
<td>3.50</td>
<td>3.28</td>
</tr>
<tr>
<td>Overall</td>
<td>4.86</td>
<td>5.05</td>
<td></td>
</tr>
</tbody>
</table>

In this study at least, bizarreness didn’t matter—overall memory performance was virtually the same whether the images were bizarre (4.86 out of 9) or not (5.05). On the other hand, combining the images produced a large effect. Memory was much better when the images were combined (6.64) than when they were not (3.28). In bar graph form, these same results are shown in Figure 8.2.

Interactions
Main effects are important outcomes in factorial designs, but the distinct advantage of factorials over single-factor designs lies in their potential to show interactive

---

1 They also tested a third factor that they called “relevance.” Some of the word pairs were accompanied by images depicting the words, as in Figure 8.1, while other word pairs were shown with irrelevant images. Recall was better in the relevant condition.
Figure 8.1 Sample stimulus materials from the imagery study by Wollen et al. (1972).

Figure 8.2 Bar graph showing a main effect for combined images, no main effect for bizarreness of the images, and no interaction (Constructed from data in Wollen et al., 1972)
Are there any main effects here? No—all the row and column means are the same. So did nothing at all happen in this study? No—something clearly happened. Specifically, the students who took the lecture course did better than those who took the lab course, but the humanities students did better than the science students. Hence, there is an interaction effect between the type of course and the discipline of the students. This is a perfect example of a single-factor design, where each factor (course type and discipline) is measured at only two levels.

<table>
<thead>
<tr>
<th>Course Type</th>
<th>Humanities</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Lab</td>
<td>80</td>
<td>70</td>
</tr>
</tbody>
</table>

In a study like this, the dependent variable would be some measure of learning. Perhaps science majors would especially benefit from the laboratory approach. To test the idea, I need to compare different types of students. Perhaps science majors and humanities majors. This calls for a 2x2 design that would look like this.

Outcomes—Main Effects and Interactions

Effects of a factorial design: an interaction is said to occur when the effect of one independent variable depends on the level of another independent variable. This is a moderately difficult concept to grasp, but it is of immense importance in understanding factorial designs. Interactions often provide the most interesting results in a factorial study. To start, consider the interaction in a study of general psychology. Perhaps science majors would especially benefit from the laboratory approach. To test the idea, I need to compare different types of students. Perhaps science majors and humanities majors. This calls for a 2x2 design that would look like this.
matching variables. In effect, you might end up with the same people who were in the factorial example. However, by running it as a single-factor design, your results would be:

Lab course: 75  Lecture course: 75

and you might conclude that it doesn’t matter whether general psychology includes a lab or not. With the factorial design, however, you know that the lab indeed matters, but only for certain types of students. In short, factorial designs can be more informative than single-factor designs. To further illustrate the concept of interactions, consider the outcome of the following case study.

Case Study 15—An Interaction with No Main Effects
There has been considerable research indicating that people remember something best if they are in the same location or context where they learned it in the first place. You might have experienced this if you were able to find your lost keys after putting yourself, either mentally or physically, back in the place where you last remembered seeing them.

The study winning the prize for the most creative test of this context-dependent learning hypothesis was carried out by Godden and Baddeley (1975). They used a $2 \times 2$ factorial in which participants learned a list of 36 words in one setting, then recalled the list in either the same setting or a different one. What makes the experiment creative is the choice of settings. Members of a diving club were the participants, and they learned the lists either on the shore of a beach or in the water at a depth of 20 feet! The first independent variable was the location where learning took place, and the two levels were “on land” and “under water.” The second variable was where recall occurred, and it also had the two levels of land and sea. Hence, there were four conditions to the study:

1. learn on land—recall on land
2. learn on land—recall underwater
3. learn underwater—recall on land
4. learn underwater—recall underwater

All divers eventually participated in all four conditions, making this a repeated-measures factorial design. The results, expressed as the average number of words recalled per list, were as follows:

<table>
<thead>
<tr>
<th>Where they learned</th>
<th>Where they recalled</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>On land</td>
<td>Underwater</td>
</tr>
<tr>
<td>On land</td>
<td>13.5</td>
<td>8.6</td>
</tr>
<tr>
<td>Underwater</td>
<td>8.4</td>
<td>14.4</td>
</tr>
<tr>
<td>Overall</td>
<td>11.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>
Combinations of Main Effects and Interactions

The example of context-dependent memory illustrates one type of outcome in a factorial design (an interaction, but no main effects), but there are many patterns...
of results that could occur. In a simple $2 \times 2$ design, for instance, there are eight
opportunities:

1. a main effect for factor A only
2. a main effect for factor B only
3. main effects for A and B only
4. a main effect for A plus an interaction
5. a main effect for B plus an interaction
6. main effects for both A and B plus an interaction
7. an interaction only, no main effects
8. no main effects, no interaction

Let's briefly consider several of these outcomes in the context of the earlier
hypothesized experiment on imagery training and presentation rate. For each of the
following examples, I have created some data that might result from the study on
the effects of imagery instructions and presentation rate on memory for a 30-word
list, translated the data into a graph, and given a verbal description of the results. I
haven't tried to create all of the eight possibilities listed above; rather, the following
elements illustrate outcomes that might actually occur in this type of study:

1. Imagery instructions improve recall, regardless of presentation rate, which
doesn't affect recall. That is, there is a main effect for factor A (imagery instruc-
tions) but no main effect for presentation rate (B). This graph should remind
you of the graph for case study 12 (Figure 8.2), except that the graph below is
a line graph instead of a bar graph.

```
<table>
<thead>
<tr>
<th>Imagery</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 sec</td>
<td>22</td>
</tr>
<tr>
<td>4 sec</td>
<td>22</td>
</tr>
<tr>
<td>Overall</td>
<td>14</td>
</tr>
</tbody>
</table>
```

2. Recall is better with slower rates of presentation, but the imagery instructions
were not effective in improving recall. That is, there is a main effect for factor B
(presentation rate) only.

```
<table>
<thead>
<tr>
<th>Imagery</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 sec</td>
<td>14</td>
</tr>
<tr>
<td>4 sec</td>
<td>22</td>
</tr>
<tr>
<td>Overall</td>
<td>18</td>
</tr>
</tbody>
</table>
```
3. Recall is better with slower rates of presentation; in addition, the imagery instructions were effective in improving recall. In this case, then, main effects for both factors occur. This is the outcome most likely to occur if you actually completed this study.

4. At the 2-second presentation rate, imagery training clearly improves recall (i.e., from 12 to 28); however, at the 4-second rate, recall is almost perfect, regardless of how subjects are trained. That is, there is an interaction between type of training and presentation rate. In this case, the interaction may have been influenced by what is called a ceiling effect, a result in which the scores for different conditions are all so close to the maximum (30 words in this example) that no difference could occur. \(^2\) Here, the imagery group recalls just about all the words, regardless of presentation rate. To test for the presence of a ceiling effect, you could replicate the study with 50-item word lists and see if performance improves for the imagery/4-second group.

You may be wondering about the obvious main effects that occur in this example. Surely the row (20 and 28) and column (also 20 and 28) means indicate significant overall effects for both factors. Technically, yes, the analysis probably would yield statistically significant main effects in this example, but this is a good illustration of the fact that interactions usually take precedence over main effects when interpreting the results. In this particular case, the main effects aren't really meaningful; the statement that imagery yields a general improvement in recall is not quite accurate. Rather, it only seems to improve recall at the faster presentation rate. Likewise, concluding that 4 seconds per

\(^2\) It is also possible for the scores in two conditions to be equal because they couldn't get any lower. Yes, these are called floor effects.
5. This is not to say that main effects never matter when an interaction exists. Consider this last example:

<table>
<thead>
<tr>
<th></th>
<th>2 sec</th>
<th>4 sec</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imagery</td>
<td>19</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>Control</td>
<td>5</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Overall</td>
<td>12</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

In this case, imagery training generally improves recall (i.e., there's a main effect for A: 21 > 10). Also, a slower presentation rate improves recall for both groups (i.e., a main effect for B also: 19 > 12). Both of these outcomes are worth reporting. What the interaction shows is that slowing the presentation rate improves recall somewhat for the imagery group (23 is a bit better than 19) but improves recall considerably for the control group (15 is a lot better than 5). Another way of describing the interaction is to say that at the fast rate, the imagery training is especially effective (19 is a lot better than 5) At the slower rate, imagery training still yields better recall, but not by as much as the slow rate (23 is somewhat better than 15).

From examining these graphs, you might have noticed a standard feature of interactions. In general, if the lines on the graph are parallel to each other, then no interaction is present. If the lines are nonparallel, however, an interaction probably exists. Of course, this rule about parallel and nonparallel lines is only a general guideline. Whether an interaction exists (in essence, whether the lines are sufficiently nonparallel) is a statistical decision, to be determined by the analysis of variance.

Identifying interactions by examining whether lines are parallel or not is easier with line graphs than with bar graphs. Hence, the guideline mentioned in Chapter 7 about line graphs being used only with continuous variables is sometimes ignored if the key finding is an interaction. For example, a study by Keltner, Ellsworth, and Edwards (1993) showed that when participants were asked to estimate the likelihood of some bad event (e.g., a car accident) occurring, there was an interaction between the emotion they were experiencing during the experiment and whether the hypothetical event was said to be caused by a person or by circumstances. When participants were feeling sad, they believed that events produced by circumstances (e.g., wet roads) were more likely to occur than events produced by individual actions (e.g., poor driving). When participants were angry, however, the opposite happened. As you can see from Figure 8.4, a line graph was drawn even though the X-axis uses a discrete variable. Keltner et al. (1993) probably wanted to show the interaction as clearly as possible, so they ignored the guideline about discrete variables. To repeat a point made earlier, when presenting any data, the overriding concern is to make one's hard-earned results as clear as possible to the reader.
actorial Designs

It's only true

In any factor-
tions should be

Interaction exists.

Before we turn to a system for categorizing the different types of factorial designs, you should read Box 8.1. It describes one of psychology's most famous experiments, a classic study supporting the idea that between the time you last study for an exam and the time you take the exam, you should be sleeping. It was completed in the early 1920s, a time when the term "factorial design" was not even in the vocabulary of experimental psychology and a time when analysis of variance, the statistical tool most frequently used to analyze factorials, was just being conceptualized. Yet the study illustrates the kind of thinking that leads to factorial studies—the desire to examine more than one independent variable at the same time.

Box 8.1

CLASSIC STUDIES—To Sleep, Perchance to Recall

As you now know, the major attraction of factorial designs is their ability to simultaneously examine the effects of two or more independent variables, as well as the interacting effects of these variables. Although the term "factorial design" and the statistical tools to analyze factorials were not used widely until after World War II (see Box 8.3), attempts to study more than one variable at a time occurred well before that time. A classic example is a study by John Jenkins and Karl Dallenbach, first reported as part of a series of studies from Cornell University in 1924. The Jenkins and Dallenbach (1924) study still appears in many general psychology books as the standard example of "retroactive interference"—the tendency for memory to be hindered if other mental activities intervene between the time of study and the time of recall. In essence, the study was a 2 x 4 repeated-measures factorial. The "2" was whether or not activities intervened between learning and a recall test, and the "4" referred to four different retention intervals: recall was tested either 1, 2, 4, or 8 hours after initial learning. What made the study interesting (and eventually, famous) was the factor with two levels. The study's participants spent the intervening time either awake and doing normal student behaviors,
or sleep in Cornell's psychology lab. The prediction that being asleep would produce larger amounts of pericentric recall was supported in both the "weakened" condition and the "enhanced" condition, and in both experiments. The difference in recall is significant and could be due to the fact that the subjects in the experimental group were more likely to have more frequent nights of solid sleep than the subjects in the control group. This suggests that sleep may have a positive impact on memory consolidation.
awake trials, the students were told when to return to the lab for recall (i.e., they knew
the retention interval), but during the asleep trials, students did not know when they
would be awakened. Jenkins and Dallenbach were aware of the problem, considered
alternatives, but decided their procedure was adequate.

The results? Figure 8.5 is a reproduction of the graph in the article, showing the data
for each student. Each data point is an average of the eight (usually) trials for each
condition of the study. Several things are clear. First, both students behaved similarly. Second,
and this was the big finding, there was a big advantage for recall after sleeping, compared
with recall after being awake. Third, there appears to be a hint of an interaction. As Jen-
kins and Dallenbach described it: “The curves of the waking experiments take the familiar
form: a sharp decline which becomes progressively flatter. The form of the curves of the
sleep experiments, however, is very different: after a small initial decline, the curves flatten
and a high and constant level is thenceforth maintained” (1924, p. 610).

There was one other intriguing outcome of the study, one never reported in textbook
accounts. As the experiment progressed, it became increasingly difficult for Dallenbach to

---

**Figure 8.5** The Jenkins and Dallenbach study on retroactive interference, showing data for both of the Cornell students who participated, L. R. Hodell (H) and J. S. McGrew (Mc). Keep in mind that the study was completed long before an ethics code would have deleted the participants’ names.

(From Jenkins & Dallenbach, 1924)
Varieties of Factorial Designs

Like the decision tree in Figure 7.1 for single-factor designs, Figure 8.6 shows you the decisions involved in arriving at one of six possible factorial designs. You’ll recognize that four of the designs mirror those in Figure 7.1, but two designs are unique to factorials. First, while the independent variable must be either a between-subjects or a within-subjects variable in a single-factor design, both types of variables can be present in a factorial design. When this happens, the design is called a mixed factorial design. In a mixed design, at least one variable must be tested between subjects, and at least one must be tested within subjects. Second, some between-subjects factors include both a manipulated independent variable and a subject variable. Because these designs can yield an interaction between the type of person in the study and the environment created in the study, they can be called P \times E factorial designs, or “Person by Environment designs,” with “environment” defined broadly to include any manipulated independent variable. A further distinction can be made within \( P \times E \) designs, depending on whether the manipulated variable is a between-subjects or a within-subjects factor. If the latter is the case, the \( P \times E \) design could be called a \( P \times E \) mixed factorial. Figures 8.7 and 8.8 show the decisions to be made when using mixed and \( P \times E \) designs. Let’s examine each of them in more detail.

Mixed Factorial Designs

In Chapter 6, you learned that when independent variables are between-subjects variables, creating equivalent groups can be a problem and procedures like random assignment can help solve the problem. Similarly, when independent variables are within-subjects variables, a difficulty arises because of potential sequence effects and counterbalancing is the normal solution. Thus, in a mixed design, the researcher usually gets to deal with both the problems of equivalent groups and the problems of sequence effects. Not always though—there is one type of mixed design where counterbalancing is not used because sequence effects are the outcome of interest. For example, in learning and memory research, “trials” is a frequently encountered within-subjects independent variable. Counterbalancing makes no sense in this case.
FIGURE 8.6 Decision tree—factorial designs.

because one purpose of the study will be to show sequential changes from trial to trial. The following two case studies show mixed designs, one requiring counterbalancing and one in which trials is the repeated measure.

Case Study 16—A Mixed Factorial with Counterbalancing
Riskind and Maddux (1993), perhaps as a consequence of seeing too many bad horror movies involving spiders, were interested in how people manage their emotions in

FIGURE 8.7 Decisions to be made with a $P \times E$ design
frightening circumstances. They created a $2 \times 2$ mixed factorial design in which they manipulated self-efficacy and "looming." Self-efficacy refers to a person's sense of competence in dealing with life's problems and is normally used in research as a subject variable. In this study, however, the feeling of self-efficacy was manipulated by the experimenters. Participants were told to visualize a situation in which they were sitting in a chair in a small room that also contained a tarantula. In the high Self-Efficacy situation, they were instructed to imagine that the door to the room was unlocked, that they were free to move around, and that they had a magazine available to swat the spider if necessary. Participants randomly assigned to the Low Self-Efficacy condition, however, were told to imagine that they were tied securely to the chair, that the door was locked, and that the newspaper was out of reach. While visualizing these circumstances, participants saw films of spiders that were either (a) stationary or moving away from them or (b) moving toward them (i.e., looming). This second variable was a repeated-measures variable, and it was presented in a counterbalanced order (unfortunately, the authors didn't report exactly which counterbalancing procedure they used). The dependent variable was a self-reported assessment of fear.

The outcome, portrayed in Figure 8.9 in both factorial matrix and graph form, is a good example of how interactions are often more important than main effects. As you can see, differences occurred for both row and column means, and both main effects were statistically significant. More important, however, was an interaction between the factors. A large amount of fear (4.50) occurred when the situation of a bad horror movie was simulated (looming spider plus low self-efficacy), while the film viewers reported only moderate to low amounts of fear in the other three conditions (2.24, 2.64, 2.73). Thus, for high self-efficacy participants, fear was fairly low regardless of the relative motion of the spider. On the other hand, for those experiencing low self-efficacy, the amount of fear was clearly affected by whether the spider was looming or not.

Case Study 17—A Mixed Factorial without Counterbalancing
A good example of the situation in which "trials" is the within-subjects factor in a mixed design is a memory procedure called "release from PI" (Wickens, Born, &
in which they son's sense of each as a sub-pulated by the they were six-high Self-Effie-the room was magazine available w' Self-Efficacy, the chair, that is visualizing these onary or mov-second variable d order (unfor-procedure they 1 graph form, is main effects. As and both main was an interac-en the situation efficacy), while the other three s, fear was fairly hard, for those ted by whether

graphic factors in aickens, Born, &

**Figure 8.9** The interactive effects of looming and self-efficacy on fear.
(From Riskind & Maddux, 1993)

Allen, 1963) PI, or proactive interference, is a phenomenon in which the learning and recall of new information is hindered by the prior learning of old information. You might have experienced this if you found it difficult to remember a new phone number because your old one kept coming to mind. The amount of interference is expected to be especially great if the old information is similar to the new information. One way to test the idea that strength of PI is a function of item similarity is to have participants study and recall a sequence of stimulus items that are similar, then switch to a different type of stimulus item. Presumably, PI should build up from trial to trial for the similar items, then "release" when the stimulus type changes. Behaviorally, this means that recall accuracy should gradually decrease while PI is accumulating, then increase again once the release occurs.

Release from PI research normally uses words or nonsense syllables as stimuli, but a study by Gunter, Berry, and Clifford (1981) took a different and more applied approach: they used items from television news shows in a series of experiments. We'll consider their experiment 1, which serves as a nice illustration of a mixed factorial design without counterbalancing.

Participants were told they would be seeing clips from a televised news broadcast and would then be tested on the content of the news items. On each trial, they saw a sequence of three stories, then worked on a distractor task (a crossword puzzle) for a minute, then tried to recall as much as they could about the stories. Each participant experienced four such trials. Half of them were randomly assigned to the release from PI condition; they went through three trials in which all the news was from a single general category (e.g., domestic political events) and a fourth ("release") trial with news from a different category (e.g., foreign political events). The remaining participants were in a control group; all four of their trials were from the same category. Thus the design was a 2 (release/no release) × 4 (trials) mixed factorial design. Whether or not participants were in the release condition was the between-subjects factor, and trials was the repeated-measures or within-subjects factor.

Figure 8.10 shows the results. That PI was operating is clear from the control group's recall scores; they steadily declined. Release from PI also occurred, as is evident from the recall scores of the experimental group.
One of the control features of this study is worth noting. A possible problem was that performance on the release trial might have gone up simply because the foreign news items were easier to recall than the domestic items. To eliminate this possibility, half of the participants in the release group saw three domestic trials followed by a foreign trial; the other half saw three foreign trials followed by a domestic trial. Likewise, in the control group, half saw four domestic trials and the remainder saw four foreign trials. The order made no difference to the results. 3

**Factorials with Subject and Manipulated Variables: P × E Designs**

Chapter 5 introduced the concept of a subject variable—some already existing attribute of an individual such as age, gender, or some personality characteristic. You also learned to be cautious about drawing conclusions when subject variables are involved. Assuming proper control, causal conclusions can be drawn with manipulated independent variables, but with subject variables such conclusions cannot be drawn. P × E designs include both subject (the “P”) and manipulated (the “E”) variables. Causal conclusions can be drawn if a significant main effect occurs for the manipulated “Environmental” factor, but they cannot be drawn when a main effect occurs for the subject variable or “Person” factor, and they also cannot be drawn if an interaction occurs. Despite this limitation, designs including both subject and manipulated variables are popular, in part because they combine the two research traditions identified by Woodworth in his famous “Columbia bible” (see the opening paragraphs of Chapter 5). The correlational tradition is associated with the study of individual differences, and the subject variable or the “P” factor in the P × E design looks specifically at these differences. A significant main effect for this factor shows that two different types of individuals perform differently on whatever behavior is being measured as the dependent variable. The experimental tradition, on the other hand, is concerned with identifying general laws of behavior that apply to

---

3 Calling the design a 2 × 4 design emphasizes the two important variables, but technically, this was a 2 × 2 × 4 mixed design, with the second “2” being the between-subjects factor of news category, foreign or domestic.
some degree to everyone, regardless of individual differences. Hence, finding a significant main effect for the manipulated or the “E” factor in a P x E design indicates that the situational factor is powerful enough to influence the behavior of many different kinds of persons. Consider a hypothetical example that compares introverts and extroverts (the “P” variable) and asks participants to solve problems in either a small, crowded room or a large, uncrowded room (the “E” variable). Suppose you get results like this (DV = number of problems solved):

<table>
<thead>
<tr>
<th></th>
<th>Small room</th>
<th>Large room</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introverts</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Extroverts</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

In this case, there would be a main effect for personality type, no main effect for environment, and no interaction. Extroverts clearly outperformed introverts (18 > 12), regardless of whether they worked in crowded conditions or not. The researcher would have discovered an important way in which individuals differ, and the differences extend to more than one kind of environment.

A very different conclusion would be drawn from this outcome:

<table>
<thead>
<tr>
<th></th>
<th>Small room</th>
<th>Large room</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introverts</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Extroverts</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

This yields a main effect for the environmental factor, no main effect for personality type, and no interaction. Now it is the environment (room size) that produced the powerful effect (18 > 12), and this effect extended beyond just a single type of individual; regardless of personality type, performance deteriorated under crowded conditions. Thus, finding a significant main effect for the “P” factor indicates that powerful personality differences occur, while finding a significant main effect for the “E” factor shows the power of some environmental influence to go beyond just one type of person. Of course, another result could be two main effects, indicating that each factor is important.

The most interesting outcome of a P x E design, however, is an interaction. When this occurs, it shows that for one type of individual, changes in the environment have one kind of effect, while for another type of individual, the same environmental changes have a different effect. Staying with the introvert/extrovert example, suppose this happened:

<table>
<thead>
<tr>
<th></th>
<th>Small room</th>
<th>Large room</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introverts</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Extroverts</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

In this case, neither main effect would be significant, but an interaction clearly occurred. One thing happened for introverts, and something different occurred for
extroverts. Specifically, introverts performed much better in the small than in the large room, while extroverts did much better in the large room than in the small one.

Factorial designs that include both subject variables and manipulated variables are popular in educational research and in research on the effectiveness of psychotherapy (Smith & Sechrest, 1991). In both areas, the importance of finding significant interactions is indicated by the fact that such designs are sometimes called ATI designs, or “Aptitude-Treatment Interaction designs.” As you might guess, the “aptitude” refers to the nonmanipulated subject (person) variable and the “treatment” refers to the manipulated, environmental variable. An example from psychotherapy research is a study by Abramovitz, Abramovitz, Roback, and Jackson (1974). Their “P” variable was locus of control. Those with an external locus generally believe that external events exert control over their lives, while internals believe that what happens to them is a consequence of their own decisions and actions. In the study, externals did well in therapy that was more directive in providing guidance for them, but they did poorly in nondirective therapy, which places more responsibility for progress on the client. For internals, the opposite was true: They did best in the nondirective therapy and not too well in directive therapy.

ATIs in educational research usually occur when the aptitude or person factor is a learning style variable and the treatment or environmental factor is some aspect of instruction. For example, Figure 8.11 shows the outcome of some educational research reported by Valerie Shute, a leading authority on ATI designs (Shute, 1994). The study compared two educational strategies for teaching basic principles of electricity: rule induction and rule application. Participants were randomly assigned to one strategy or the other. The subject variable was whether learners scored high or low on a measure of “exploratory” behavior. The graph shows that those scoring high on exploratory behavior performed better in a rule induction setting, where

![Graph showing aptitude-treatment interaction](image)

**Figure 8.11** A P × E interaction between level of exploratory behavior and type of learning environment

(From Shute, 1994)
Varieties of Factorial Designs

they were asked to do more work on their own, while those scoring low on exploratory behavior performed better in a rule application environment, where the educational procedures were more structured for them.

P × E factorial designs are also very popular in personality research and in health psychology. The following case study is a good example of a study from health psychology that also examines a personality factor.

Case Study 18—A Factorial Design with a P × E Interaction

Over the past 25 years or so, there has been considerable interest in the personality features of those with a so-called Type A behavior pattern (Friedman & Rosenman, 1974) Type A individuals are competitive, achievement oriented, and compulsive about many things, including time (they are often early, never late). They try to do many things at once and seem to have a very high energy level. More ominously, Type A's who combine these traits with a generally hostile attitude tend to develop coronary heart disease. That is, under certain circumstances, a Type A behavior pattern can be decidedly unhealthy.

Research on Type A behavior often uses a P × E factorial design. The “P” (subject) variable is the behavior pattern—Type A or its antithesis, the more laid-back Type B. Selection of participants for each of these groups is determined by scores on some test for the Type A/B patterns. The “E” variable is some factor manipulated by the experimenter. For example, in a study by Holmes, McGilley, and Houston (1984), Type A and B college students were differentiated by their scores on a student version of the Jenkins Activity Survey, a test frequently used in research on Type A behavior patterns. A total of 394 students took the test, and the researchers recruited 30 high scorers and 30 low scorers. Participants in these two groups then were randomly assigned to one of three tasks that differed in how challenging they were. Hence the manipulated independent variable was task difficulty. The task was a “digit span” procedure, borrowed from an IQ test, in which participants listened to a sequence of numbers (e.g., 3-4-8) and had to repeat them in reverse order (8-4-3). The three levels of difficulty were defined in terms of the number of digits read to participants during the six trials completed by each of them: two, five, or seven. Arousal during the task, operationally defined in terms of several physiological measures, including systolic blood pressure, was the dependent variable.

Figure 8.12 shows the P × E interaction that occurred. Arousal for Type A’s and B’s did not differ on easy and moderate tasks, but on very difficult tasks, systolic pressure continued to increase for the A’s but leveled off for the B’s. That is, compared to Type B’s, Type A’s showed significantly elevated systolic blood pressure, but only on really challenging tasks. There was no overall main effect for personality type, but there was a main effect for task difficulty, as you can detect from the generally increasing blood pressure as the task became more difficult. In sum, the study showed that “differences in arousal between Type A and Type B persons are most likely to emerge at high levels of challenge” (Holmes et al., 1984, p. 1326), a finding consistent with the highly competitive, achievement focus of Type A persons.

By the way, those advocating the use of P × E designs can trace their partiality to the work of Kurt Lewin (1890–1947), a pioneer in social and child psychology. The central theme guiding Lewin’s work was that a full understanding of behavior
required studying both the person's attributes and the environment in which the person operated. He expressed this idea in terms of the well-known formula, $B = f(P, E)$—behavior is a joint function of the person and the environment (Goodwin, 1999). P x E factorial designs, which derive their name from Lewin's formula, are perfectly suited for discovering the kinds of interactive relationships that Lewin believed characterized human behavior.\(^4\)

**Recruiting Participants for Factorial Designs**

It should be evident from the definitions of the varieties of factorials that the number of participants needed could vary considerably, depending on the design. If you need 5 participants to fill one of the cells in the 2 x 2 factorial, for example, the total number of people to be recruited could be 5, 10, or 20. Figure 8.13 shows why. In Figure 8.13a, both variables are tested between subjects and 5 different participants will be needed per cell, for a total of 20. In Figure 8.13b, both variables are tested within subjects, making the design a repeated-measures factorial. The same five individuals will contribute data to each of the four cells. In a mixed design, Figure 8.13c, one of the variables is tested between subjects and the other is tested within subjects. Thus, 5 participants will participate in two cells and 5 will participate in the other two cells, a total of 10.

Knowing how many participants to recruit for an experimental leads naturally to the question of how to treat the people who arrive at your experiment. Box 8.2 provides a hands-on, practical guide to being an ethical researcher.

---

\(^4\) One unfortunate implication of Lewin's choice of the label "P" is the implication that a P x E design involves human participants only. Yet it is quite common for such a design to be used with animal subjects (a study in which the subject variable is the gender of the primates being tested for instance).
Varieties of Factorial Designs

(a) For a $2 \times 2$ design with four different groups and five participants per cell—20 subjects needed

| S1   | S11 |
| S2   | S12 |
| S3   | S13 |
| S4   | S14 |
| S5   | S15 |
| S6   | S16 |
| S7   | S17 |
| S8   | S18 |
| S9   | S19 |
| S10  | S20 |

(b) For a $2 \times 2$ repeated—measures design with five participants per cell—five subjects needed

| S1   | S1 |
| S2   | S2 |
| S3   | S3 |
| S4   | S4 |
| S5   | S5 |
| S1   | S1 |
| S2   | S2 |
| S3   | S3 |
| S4   | S4 |
| S5   | S5 |

(c) For a $2 \times 2$ mixed design with five participants per cell—10 subjects needed

| S1   | S1 |
| S2   | S2 |
| S3   | S3 |
| S4   | S4 |
| S5   | S5 |
| S6   | S6 |
| S7   | S7 |
| S8   | S8 |
| S9   | S9 |
| S10  | S10 |

**Figure 8.13** Participant requirements in factorial designs.
Box 8.2

ETHICS—On Being a Competent and Ethical Researcher

You learned about the APA code of ethics in Chapter 2, and you have been encountering Ethics boxes in each of the subsequent chapters. Although you should have a pretty good sense of what the ethical requirements of a study are (consent, debriefing, etc.), you might not be quite sure how to put this into actual practice. Hence, this might be a good time to give you a list of practical tips for being an ethically responsible experimenter. Here goes.

✓ Don’t just get to your session on time—be early enough to have all the materials organized and ready to go when your participants arrive.

✓ Always treat the people who volunteer for your study with the same courtesy and respect that you would hope to receive if the roles were reversed. Greet them when they show up at the lab and thank them for signing up and coming to the session. They might be a bit apprehensive about what will happen to them in a “psychology” experiment, so your first task is to put them at ease, while at the same time maintaining your professional role as the person in charge of the session. Always remember that they are doing you a favor—the reverse is not true. Smile often.

✓ Start the session with the informed consent form and don’t convey the attitude that this is a time-consuming technicality that has to be completed before the important part starts. Instead, make it clear that you want your participants to have a clear idea of what they are about to do. If they don’t ask questions while reading the consent form, be sure to ask them if they have any questions when they finish reading. Make sure there are two copies of the signed consent form—one for them to take with them and one for your records.

✓ It is a good idea to write out the “Instructions to Participants” ahead of time. Depending on the study, you can read them the instructions, or give them the instructions to read. If it isn’t necessary to have elaborate written instructions, at least have a list with you that includes the key points of the instructions about the procedure.

✓ Before you test any “real” participants, practice playing the role of experimenter a few times with friends or lab partners. Go through the whole experimental procedure—think of it as a
dress rehearsal and an opportunity to iron out any problems with the procedure.

☑ Be alert to any signs of distress in participants during the session. Depending on the constraints of the procedure, this could mean halting the study and discarding their data, but their welfare is more important than your data. Also, you're not a professional counselor—if they seem disturbed by their participation, gently refer them to your course instructor.

☑ Prepare the debriefing carefully. As a student experimenter, you probably won't be running studies involving elaborate deceptions or producing high levels of stress, but you will be responsible for making this an educational experience for your participants. Hence, you should work hard on a simplified description of what the study hopes to discover and you should give them the chance to suggest improvements in the procedure or ideas for the next study. So don't rush the debriefing or give a cursory description that gives the impression that you hope they will just leave. And if they seem to want to leave without any debriefing (some will), don't let them. Debriefing is an important part of your responsibility as a researcher and an educator. (Of course, if they say, "I thought you said we could leave any time," there's not much you can do!)

☑ Before they go, remind them that the information on the consent form gives them names of people to contact about the study. Give them a rough idea of when the study will be completed and when they can expect to hear about the overall results. Also, ask them not to discuss the experiment with others who might be participants. Leakage, a tendency for participants to tell others about the studies they've completed, can be a serious problem, especially at small schools (see Box 6.3, p. 208, for more on the responsibilities of research participants). If you have been good to them throughout the session, however, that increases the chances of their cooperation in this regard.

☑ As they are leaving, be sure to thank them for their time and effort and be sure you are smiling as they go out the door. Remember that some of the students you test will be undecided about a major and perhaps thinking about psychology; their participation in your study could enhance their interest.

Analyzing Factorial Designs

We've already seen that multilevel, single-factor designs are analyzed with a 1-way ANOVA. Similarly, factorial designs using interval or ratio data are analyzed with N-way ANOVAs, with N referring to the number of independent variables
involved. Hence, a $2 \times 3$ factorial would be analyzed by a 2-way ANOVA and a $2 \times 2 \times 4$ by a 3-way ANOVA.

When doing a 1-way ANOVA, just one $F$-ratio is calculated. Then there may be subsequent testing if the $F$ is significant. For a factorial design, however, more than one $F$-ratio will be calculated. Specifically, there will be an $F$ for each possible main effect and for each possible interaction. For example, in the $2 \times 2$ design investigating the effects of imagery training and presentation rate on memory, an $F$-ratio will be calculated to examine the possibility of a main effect for type of training, another for the main effect of presentation rate, and a third for the potential interaction between the two (see Appendix C for an example of a $2 \times 2$ ANOVA). In an $A \times B \times C$ factorial, there will be seven $F$-ratios calculated: three for the main effects of $A$, $B$, and $C$, three more for the 2-way interaction effects of $A \times B$, $B \times C$, and $A \times C$, plus one for the 3-way interaction, $A \times B \times C$. Subsequent testing may occur with factorial ANOVAs as well as with 1-way ANOVAs. For example, in a $2 \times 3$ ANOVA, a significant $F$ for the factor with three levels would trigger a subsequent analysis (e.g., Tukey's HSD—see Appendix C) that compared the overall performance of levels 1 and 2, 1 and 3, and 2 and 3.

Before closing this chapter, let me make one final point about factorials and the analysis of variance. You've been looking at many factorial matrices in this chapter. They might vaguely remind you of farming. If so, it's no accident, as you can discover by reading Box 8.3, which tells you a bit about Sir Ronald Fisher, who invented the analysis of variance.

---

**Box 8.3**

**ORIGINS—Factorials Down on the Farm**

Imagine that you're in a small plane flying over Kansas. Looking out the window, you see mile after mile of farms, with their fields laid out in blocks. The pattern might remind you of the $2 \times 2$ and $3 \times 3$ factorial matrices that you've just encountered in this chapter. This is because factorial designs and the ANOVA procedures for analyzing them were first developed in the context of agricultural research, devised by Sir Ronald Fisher. The empirical question was, “What are the best possible conditions or combinations of conditions for raising crop X?”

Ronald Aylmer Fisher (1890–1962) was one of Great Britain's best-known statisticians, equal in rank to the great Karl Pearson, who invented the correlation measure we now call “Pearson's $r$” (next chapter). Fisher created statistical procedures useful in testing predictions about genetics, but he is perhaps best known among research psychologists for creating the analysis of variance, which yields the $F$-ratios that allow one to decide about the null hypothesis in experimental research. You can easily guess what the “F” represents.
For about a 15-year period beginning in 1920, Fisher worked at an experimental agricultural station at Rothamsted, England. While there, he was involved in research investigating the effects on crop yield of such variables as fertilizer type, rainfall level, different planting sequences, and different genetic strains of various crops. He published articles with titles like "Studies in Crop Variation. VI. Experiments on the Response of the Potato to Potash and Nitrogen" (Kendall, 1970, p. 447). In the process, he invented analysis of variance as a way of analyzing the data. He emphasized the importance of using factorial designs, “for with separate [single-factor] experiments we should obtain no light whatever on the possible interactions of the different ingredients” (Fisher, 1935/1951, p. 95, italics added). In the real world of agriculture, crop yields resulted from complex combinations of factors, and studying one factor at a time wouldn’t allow a thorough evaluation of those interactive effects.

A simple 2 x 2 design for one of Fisher’s experiments, with each block representing how a small square of land was treated, might look like Figure 8.14. As with any factorial, this design allows one to evaluate main effects (of fertilizer and type of wheat in this case), as well as the interaction of the two factors. In the example in Figure 8.14, if we assume that the shaded field produces significantly more wheat than the other three (which equal each other), then we would say that an interaction clearly occurred—the fertilizer was effective, but only for one specific strain of wheat.

Fisher first published his work on ANOVA in book form in 1925 (a year after Jenkins and Allenden published their classic sleep and memory study), as part of a larger text on statistics (Fisher, 1925). His most famous work on ANOVA, which combined a discussion of statistics and research methodology, appeared 10 years later as *The Design of Experiments* (Fisher, 1935/1951). ANOVA techniques and factorial designs were slow to catch on in the United States, but by the early 1950s, they had become institutionalized as a dominant statistical tool for experimental psychologists (Rucci & Tveney, 1980).

This completes our two-chapter sequence about experimental designs. It’s a pair of chapters (along with Chapters 5 and 6) sure to require more than one reading and much practice with designs before you’ll feel confident about your ability to use “experimental psychologist language” fluently and to create a methodologically sound experiment that is a good test of your hypothesis. Next up is a closer look at another research tradition, in which the emphasis is not on examining differences, but in examining degrees of association between measured variables.
Chapter Summary

Factorial Essentials
Factorial designs examine the effects of more than one independent variable. Factorials are identified with a notation system that identifies the number of independent variables, the number of levels of each variable, and the total number of conditions. For example, a $2 \times 3$ ("2 by 3") factorial has two independent variables, the first with two levels and the second with three levels, and six different conditions ($2 \times 3$).

Outcomes—Main Effects and Interactions
The overall influence of an independent variable in a factorial study is called a main effect. There are two possible main effects in a $2 \times 3$ design, one for the factor with two levels and one for the factor with three levels. The main advantage of a factorial design over studies with a single independent variable is that factorials allow for the possibility of discovering interactions between the factors. In an interaction, the influence of one independent variable differs for the different levels of the other independent variable. The outcomes of factorial studies can include significant main effects, interactions, both, or neither.

Varieties of Factorial Designs
All of the independent variables in a factorial design can be between-subjects factors or all can be within-subjects factors. Between-subjects factorial designs can include independent groups, matched groups, or nonequivalent groups. Within-subjects factorials are also called repeated-measures factorial designs. A mixed factorial design includes at least one factor of each type (between and within). Factorials with at least one manipulated variable and at least one subject variable allow for the discovery of Person × Environment ($P \times E$) interactions. When these interactions occur, they show how stimulus situations affect one type of person one way and a second type of person another way. A main effect for the $P$ factor (i.e., subject variable) indicates important differences between types of individuals that exist in several environments. A main effect for the $E$ factor (i.e., manipulated variable) indicates important environmental influences that exist for several types of persons. In educational research and research on the effectiveness of psychotherapy, these interactions between persons and environments are sometimes called Aptitude-Treatment-Interactions (ATIs).

Chapter Review

Multiple Choice

1. A researcher predicts that introverts will do better on a problem-solving task if they do it by themselves rather than in front of an audience. Extroverts, however, are expected to do better with an audience than alone. The researcher
   a. is predicting a main effect for the audience variable
   b. is using a mixed factorial design
Chapter Review

2. A $3 \times 3$ mixed factorial design uses five people in cell A1B1. How many people need to be recruited for this study?
   a. 5
   b. 10
   c. 45
   d. 15

3. A $2 \times 3 \times 5$ factorial design has
   a. 10 different independent variables
   b. 25 different conditions
   c. 3 different independent variables
   d. a maximum of 30 main effects

4. A $2 \times 2$ mixed factorial design will always include
   a. two subject variables
   b. a between-subjects factor and a within-subjects factor
   c. a repeated-measures variable
   d. one manipulated variable and one subject variable

5. In a maze-learning experiment, 30 rats are tested with the lights on and 30 more with the lights off. Also, within each of the two groups, 10 rats receive reinforcement as soon as they reach the goal, 10 others are reinforced 5 seconds after they reach the goal, and the remaining 10 are fed 10 seconds after reaching the goal. What can you say about this design?
   a. both independent variables are subject factors
   b. it is best described as a $2 \times 2$ independent groups design
   c. it is best described as a $2 \times 3$ mixed design
   d. six different conditions are being tested

Short Essay

1. In the context of a factorial design, distinguish between the concepts of “levels” and “conditions.”

2. What is meant by a main effect? In terms of the contents of a factorial matrix, how does one go about determining if a main effect has occurred?

3. Use the Godden and Baddeley study (Scottish diving club) to illustrate the fact that important results can occur in a study, even if no main effects occur.

4. In a study with both main effects and an interaction, explain why the interaction must be interpreted first, and the statistically significant main effects might have little meaning for the overall outcome of the study.

5. What is a ceiling effect? If one occurs, how might the method be adjusted to eliminate the effect?

6. Distinguish between a mixed factorial design and a $P \times E$ design.

7. Use the introvert/extrovert and room size example to show how $P \times E$ designs can discover important ways in which (a) individuals differ and (b) situations can be more powerful than individual differences.
8. Mixed factorial designs may or may not involve counterbalancing. Explain.
9. Describe the basic research design and the general outcome of Jenkins and Dallenbach's famous study on sleep and memory.
10. What is meant by the concept of leakage and how might it be prevented?

Applications Exercises

Exercise 8.1.—Identifying Designs

For each of the following descriptions of studies, identify the independent and dependent variables involved, the levels of the independent variable, and the nature of each independent variable (between-subjects or within-subjects; manipulated or subject variables). Identify the measurement scale for each dependent variable. Also, describe the number of independent variables and levels of each by using the factorial notation system (e.g., 2 × 3), and use Figure 8.6 to identify the experimental design.

1. On the basis of scores on the Jenkins Activity Survey, three groups of participants are identified: Type A, Type B, and intermediate. An equal number of participants in each group is given one of two tasks to perform. One of the tasks is to sit quietly in a small room and estimate, in the absence of a clock, when 2 full minutes have elapsed. The second task is to make the same estimate, except that while in the small room, the participant is playing a handheld video game.

2. College students in a cognitive mapping study are asked to use a direction finder to point accurately to three unseen locations that differ in distance from the lab. One is a nearby campus location, one is a nearby city, and the third is a distant city. Half of the participants perform the task in a windowless room with a compass indicating the direction of north. The remaining participants perform the task in the same room without a compass.

3. In a study of touch sensitivity, two-point thresholds are measured on 10 different skin locations for an equal number of blind and sighted adults. Half of the participants perform the task in the morning and half in the evening.

4. Three groups of preschoolers are put into a study of delay of gratification in which the size of the delay is varied. Children in all three groups complete a puzzle task. One group is told that as payment they can have a dollar now or three dollars tomorrow. The second group chooses between a dollar now or three dollars two days from now, and the third group chooses between a dollar now or three dollars three days from now. For each of the three groups, half the children solve an easy puzzle and half solve a difficult puzzle. The groups are formed in such a way that the average parents' income is the same for children in each group.

5. In a study of visual illusions and size perception, participants adjust a dial that alters one of two stimuli. The goal is to make the two stimuli appear to be equal in size, and the size of the error in this judgment is measured on each trial. Each participant completes 40 trials. On half of the trials, the pairs of stimuli are in color; on the other half, they are in black and white. For both the colored and
the black-and-white stimuli, half are presented at a distance of 10 feet from the participant and half are presented 20 feet away.

6. In a study of reading comprehension, sixth-grade students read a short story about baseball. The students are divided into two groups based on their knowledge of baseball. Within each group, half of the students are high scorers on a test of verbal IQ, while the remaining students are low scorers.

Exercise 8.2.—Main Effects and Interactions

For each of the following studies:

a. identify the independent variables, and the levels of each, and the dependent variable
b. place the data into the correct cells of a factorial matrix
c. determine if main effects and/or interactions exist
d. give a verbal description of the study's outcome
e. draw a graph of the results

For the purposes of the exercise, assume that a difference of “2” between any of the row or column or cell means is a significant difference.

1. A researcher is interested in the effects of ambiguity and number of bystanders on helping behavior. Participants fill out a questionnaire in a room with zero or two other people who appear to be other subjects but aren’t. The experimenter distributes the questionnaire and then goes into the room next door. After 5 minutes there is a loud crash, possibly caused by the experimenter falling. For half of the participants, the experimenter unambiguously calls out that he has fallen, is hurt, and needs help. For the remaining participants, the situation is more ambiguous—the experimenter says nothing after the apparent fall. Twenty participants are tested in each condition, and the experimenter records how long it takes (in seconds) before a participant offers help.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Mean Time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 bystanders, ambiguous</td>
<td>24</td>
</tr>
<tr>
<td>2 bystanders, ambiguous</td>
<td>38</td>
</tr>
<tr>
<td>0 bystanders, unambiguous</td>
<td>14</td>
</tr>
<tr>
<td>2 bystanders, unambiguous</td>
<td>14</td>
</tr>
</tbody>
</table>

2. A researcher interested in maze learning hypothesizes that maze learning can be influenced by the size of the reinforcer and by the delay in reinforcement. Animals are given 40 learning trials in a maze, and the number of errors per trial is recorded. There are six conditions. At the end of the maze there is

a. a 10-mg food pellet, and it is given to the animal immediately, or

b. a 10-mg food pellet, but the animal has to wait 10 seconds in the goal box before the food is given, or

c. a 10-mg food pellet, but the animal has to wait 20 seconds in the goal box before the food is given, or

d. a 20-mg food pellet, and it is given to the animal immediately, or
e. a 20-mg food pellet, but the animal has to wait 10 seconds in the goal box before the food is given, or
f. a 20-mg food pellet, but the animal has to wait 20 seconds in the goal box before the food is given

The average number of errors for animals in the six groups are:

1: 10  2: 40  3: 45  4: 5  5: 35  6: 40

Exercise 8.3.—Estimating Participant Needs

For each of the following, use the available information to determine how many research participants will be needed to complete the study (hint: one of these is unanswerable without more information):

1. a 3 × 3 mixed factorial; cell A1B1 needs 10 participants.
2. a 2 × 3 repeated-measures factorial; cell A1B1 needs 20 participants.
3. a 2 × 2 × 2 independent groups factorial; cell A1B1C1 needs 5 participants.
4. a 2 × 4 mixed factorial; cell A1B1 needs 8 participants.