Chapter 9: Factorial Designs

So far, we have focused our consideration of experimental design on the simplest possible designs with a single independent variable with just two conditions administered either between or within groups of participants. Most studies in psychology are more complex than this and, in this chapter, we start to discuss slightly more complex designs. As the complexity of the experimental design increases, the relationship of the data to the experimental hypotheses also increases. These designs allow us to test more interesting and complicated ideas about how psychological constructs interact with each other. However, these designs make the process of drawing inferences from experimental data more challenging. In this chapter, we will be concerned with how to design an experiment with multiple factors (independent variables). In Chapter 10, we will review the process of evaluating various patterns of data that can arise from these designs and how we draw conclusions from these.

## Cleanliness and moral judgments

Simone Schnall and her colleagues carried out a series of simple studies examining an interesting effect where, in which they found that priming the idea of cleanliness (Experiment 1) or washing one’s hands (Experiment 2) led people to view moral transgressions as less wrong (Schnall, Benton, Harvey, 2008). In a separate study, Schnall and her colleagues investigated whether feeling physically disgusted causes people to make harsher moral judgments (Schnall, Haidt, Clore, Jordan, 2008). In this experiment, they extended this idea to both include disgust created by the testing environment but also accounting for differences in the participants’ sensitivity to their own bodily sensations. Participants’ feelings of disgust were manipulated by testing them in either a clean room or a messy room that contained dirty dishes, an overflowing wastebasket, and a chewed-up pen. In addition, a self-report questionnaire to measure the amount of attention that people pay to their bodily sensation, described as “private body consciousness.” The primary dependent variable remained the same as in the previous simpler 2-group designs. They measured the harshness of people’s moral judgments by describing different behaviors (e.g., eating one’s dead dog, failing to return a found wallet) and having participants rate the moral acceptability of each one on a scale of 1 to 7. The primary results of this study were that participants in the messy room were, in fact, more disgusted and made harsher moral judgments than participants in the clean room—but only if they scored relatively high in private body consciousness.

The conclusion drawn in this study depends on describing an interaction between two different variables that were hypothesized to affect the dependent variable (moral acceptability rating). To be able to say that environmentally elicited disgust affected moral judgments but only for people with high sensitivity to their bodily sensations requires a **factorial design** to incorporate both variables simultaneously. With information about both the environmental variable and the participant variable, the researchers could consider three hypotheses simultaneously. First, did the messiness of the room by itself affect the moral judgments? Second, did participants who scored higher in private body consciousness rate moral judgments differently than those who scored lower? And third, did the messy room affect the higher scoring participants more than the lower scoring participants? This third hypothesis is based on an **interaction** among the variables. The main conclusion of the study is actually focused on this interaction and looking for these interactions is the primary reason to employ factorial designs in experimental research.

Factorial designs depend on all the same basic experimental design elements discussed in previous chapters, including operational definitions of psychological constructs and control of extraneous variables. The key difference is the use of multiple independent variables that are manipulated (or measured) simultaneously. Studies with multiple dependent variables are also possible, but “multivariate” research design is beyond the scope of this introductory research methods text.

# Design of Factorial Experiments

## Learning Objectives

1. Understanding one factor designs with more than two levels of the independent variable
2. Explain why researchers often include multiple independent variables in their studies.
3. Define factorial design and use a factorial design table to represent and interpret simple factorial designs.
4. Understand the core hypotheses embedded in a factorial design: main effects and interactions among effects

## One Factor Design

The simplest extension from the designs we have discussed so far with 2 groups or conditions is to consider an experimental design with three different options for the independent variable. For more complicated designs, the term **factor** is often used instead of or synonymously with the term **independent variable**. The conditions that are implemented are described as **levels** of the factor.

As an example, consider a hypothetical design where participants listed to one of three kinds of auditory input while performing a spatial cognition task. The type of music is the experimental factor and the three levels are classical music, electronic dance music, and soothing ocean sounds. Participants completed as many problems as they could in 10 minutes on a “paper cutting and folding” test. Just as in prior designs, the hypothesis is that the type of sounds listened to would influence the score on the test. However, it should be clear that there are already more outcomes to consider than we would have with a 2-group design. With a 2-group design, either the independent variable affects the dependent variable (test scores) or it does not. With three groups, the null hypothesis is that the sounds have no effect and that all three conditions are essentially identical. But we can reject that null hypothesis if any of the 3 groups shows different performance on the test from the others. The statistical tool to carry out this type of inferential statistic is the Analysis of Variance (ANOVA) which will be discussed in Chapter 10. This analysis provides a p-value that indicates the probability of the data occurring under the null hypothesis and if less than .05, we can conclude the different levels of the factor affected the dependent variable score.

In general, that is only the first step in analyzing data with more than 2 levels. We typically want to know not just that performance differs, but which of the conditions differ from each other. The statistical tool used here is the **post-hoc t-test** to do all the possible pairwise comparisons and find the differences. Conceptually, what we want to know is (1) “did classical music lead to different performance than ocean sounds?” (2) “did classical music lead to different performance than electronic dance music?” and (3) "did electronic dance music lead to different performance than ocean sounds?” Each of those potential conclusions may have very different meanings for a theory of how auditory input affects spatial cognitive performance. The first thing to note is that this simple extension to just 3 conditions instead of 2 requires us to bring in a new statistical tool, ANOVA, and do a total of 4 analyses to try to understand our data.

In practice, the challenge of drawing inferences from the data in these designs can be even harder when the data are messy. For example, we might observe that performance during classical music is reliably better than ocean sounds, but neither of the other two comparisons is statistically reliable, e.g. performance during the dance music is in-between ocean sounds and classical. These data would leave us in a difficult position for summarizing the findings of our study because different kinds of music both are and are not affecting the dependent variable.

The potential for problem in getting a strong conclusion from this kind of design makes factors with many levels less common in psychological research than just using two levels. Very complex designs with many levels on the factors do get used but often in specific cases with very strong theoretical foundations and in conjunction with more complex analytical tools. All the statistical tools that will be described in this class are simplified cases derived from a more general approach based on **general linear models**. Extrapolating to these more complex types of analysis is beyond the scope of this text.

## Factorial design

A very common approach in psychological science is to design studies with more than one factor. Researchers’ inclusion of multiple independent variables in one experiment is further illustrated by the following actual titles from various professional journals:

The Effect of Age and Divided Attention on Spontaneous Recognition

The Effects of Temporal Delay and Orientation on Haptic Object Recognition

Opening Closed Minds: The Combined Effects of Intergroup Contact and Need for Closure on Prejudice

Effects of Expectancies and Coping on Pain-Induced Intentions to Smoke

The Effects of Reduced Food Size and Package Size on the Consumption Behavior of Restrained and Unrestrained Eaters

In each of these cases, we see research that is assessing the effect of at least two factors (independent variables) on some behavior of interest. In each of these studies, the researchers are looking simultaneously at two different IV’s that may affect the DV measure. This approach goes importantly beyond examining the effect of each factor by also allowing the researchers to identify interactions between these variables that could not be assessed by doing two successive studies looking at each factor in isolation. Taking the first headline above as an example, we might find that divided attention leads to worse performance on spontaneous recognition but also that this effect is much larger for older adults than younger adults. This would be an example of an **interaction** among the experimental factors. These are generally the most interesting effects to study in psychological research but also ones that can pose more difficulties in drawing accurate inferences from.

By far the most common approach to including multiple factors (independent variables) in an experiment is the **factorial design**, which assesses both the effects of these factors and their interactions. In a factorial design, each level of one independent variable is combined with each level of the others to produce all possible combinations. Each combination, then, becomes a condition in the experiment. Imagine, for example, an experiment on the effect of cell phone use (yes vs. no) and time of day (day vs. night) on driving ability. This is shown in the factorial design table below. The columns of the table represent cell phone use, and the rows represent time of day. The four cells of the table represent the four possible combinations or conditions: using a cell phone during the day, not using a cell phone during the day, using a cell phone at night, and not using a cell phone at night. This particular design is referred to as a 2 × 2 (read “two-by-two”) factorial design because it combines two variables, each of which has two levels.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Cell Phone | |
| No | Yes |
| Time of Day | Daytime |  |  |
| Nighttime |  |  |

If one of the independent variables had a third level (e.g., using a handheld cell phone, using a hands-free cell phone, and not using a cell phone), then it would be a 3 × 2 factorial design, and there would be six distinct conditions. Notice that the number of possible conditions is the product of the numbers of levels. A 2 × 2 factorial design has four conditions, a 3 × 2 factorial design has six conditions, a 4 × 5 factorial design would have 20 conditions, and so on. Also notice that each number in the notation represents one factor, one independent variable. So by looking at how many numbers are in the notation, you can determine how many independent variables there are in the experiment. 2 x 2, 3 x 3, and 2 x 3 designs all have two numbers in the notation and therefore all have two independent variables. The numerical value of each of the numbers represents the number of levels of each independent variable. A 2 means that the independent variable has two levels, a 3 means that the independent variable has three levels, a 4 means it has four levels, etc. To illustrate a 3 x 3 design has two independent variables, each with three levels (9 conditions), while a 2 x 2 x 2 design has three independent variables, each with two levels (8 conditions). As noted in the discussion of one-factor designs, having 3 levels adds surprising amounts of complexity to interpretation. As a result, it is more common to extend designs to additional factors such as a 2 x 2 x2 design.

In principle, factorial designs can include any number of independent variables with any number of levels. For example, an experiment could include the type of psychotherapy (cognitive vs. behavioral), the length of the psychotherapy (2 weeks vs. 2 months), and the sex of the psychotherapist (female vs. male). This would be a 2 × 2 × 2 factorial design and would have eight conditions. The table below shows one way to diagram this design. In practice, it is unusual for there to be more than three independent variables with more than two or three levels each. This is for at least two reasons: For one, the number of conditions can quickly become unmanageable. For example, adding a fourth independent variable with three levels (e.g., therapist experience: low vs. medium vs. high) to the current example would make it a 2 × 2 × 2 × 3 factorial design with 24 distinct conditions. Second, the number of participants required to populate all of these conditions (while maintaining a reasonable ability to detect a real underlying effect) can render the design unfeasible (for more information, see the discussion about the importance of adequate statistical power in Chapter 11). As a result, we will primarily focus on designs with two independent variables. The general principles discussed here extend in a straightforward way to more complex factorial designs.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Psychotherapy Type | |
| Cognitive | Behavioral |
| Length | Two weeks | Therapist  Female Male | Therapist  Female Male |
| Two months | Therapist  Female Male | Therapist  Female Male |

## Assigning Participants to Conditions

The diagrams in the preceding section are useful in experimental design for planning how to assign participants to conditions and planning the total number of participants to be enrolled in the study. Recall that in a simple between-participants design, each participant is tested in only one condition. In a simple within-participants design, each participant is tested in all conditions. In a factorial experiment, the decision to take the between-participants or within-participants approach must be made separately for each independent variable. In a between-participants factorial design, all of the independent variables are manipulated between participants. For example, each participant would be tested either while using a cell phone or while not using a cell phone and either during the day or during the night. This would mean that each participant would be tested in one and only one of the four possible conditions. This type of design avoids any possible problems with order effects but does generally require a lot of participants to be recruited and enrolled in the study. In modern psychological studies, we prefer having 20-30 participants in each of the conditions meaning a 2 x 2 design might require 80-120 participants.

It's perfectly acceptable to organize the design as an entirely **within-participants factorial design** with all of the independent variables are manipulated within participants. In this case, all participants are tested in all four of the conditions, that is, each participant is tested both while using a cell phone and while not using a cell phone and both during the day and during the night. The advantages and disadvantages of these two approaches are the same as those discussed in Chapter 7. The between-participants design is conceptually simpler, avoids order/carryover effects, and minimizes the time and effort of each participant. The within-participants design is more efficient for the researcher and controls extraneous participant variables.

Since factorial designs have more than one independent variable, it is also possible to manipulate one independent variable between participants and another within participants. This is called a mixed factorial design. For example, a researcher might choose to treat cell phone use as a within-participants factor by testing the same participants both while using a cell phone and while not using a cell phone (while counterbalancing the order of these two conditions). But they might choose to treat time of day as a between-participants factor by testing each participant either during the day or during the night (perhaps because this only requires them to come in for testing once). Thus, each participant in this mixed design would be tested in two of the four conditions.

An important difference to keep in mind across these design choices is that there are slightly different statistical tools for analyzing data when there is at least one within-participants factor. For statistical analysis, the within-participants factor is typically referred to as having **repeated measures** in the design. This changes some details of how the analytical tools are run and how the data are formatted for analysis. This will be reviewed in Chapter 11.

Regardless of whether the design is between participants, within participants, or mixed, the actual assignment of participants to conditions or orders of conditions is typically done randomly. A diagram of the design can be used to both plan the total sample size and also track the accumulation of data so that the number of participants in each condition stays relatively balanced. For statistical analysis, it is best if the number of participants in each of the design cells (conditions) is the same or similar when data collection is completed.

## Non-Manipulated Independent Variables

In many factorial designs, one of the independent variables can also be a non-manipulated independent variable. In this case, the researcher measures but does not manipulate the factor and is often a characteristic that varies across participants. The study by Schnall et al. (2008) is an example of this that incorporated the participants rating of their “private body consciousness” in the design. Scores on a measure of this characteristic were used to assign participants to either high or low “condition” on this measure. In design of this kind of factor, it is necessary to have a plan for the distinction of the rating scale into the high/low categories. This can be done by using prior research with the scale provide definitions of the categories. It can also be done by using a **median split** of participants. Since the median value in a group is defined as the number that splits the groups into two equal halves, this technique is guaranteed to give equal sized samples across the two levels of this factor.

Studies with this generally approach to design are extremely common and can provide important insight into how the manipulated independent variable might have different effects on different people. In the Schnall et al. (2008), the manipulated variable was the environment, specifically how messy the room was in which participants made moral judgments. In the process of science, it is not uncommon to have developed a hypothesis that the messy room might cause people to make harsher moral judgments, implying a typical two-condition research study that is also consistent with prior research published by the same group. However, in data collection, it might become clear that the effect of the room is not statistically reliable in the simple design, leading researchers to examine why this effect might be influencing some participants but not others. That might provide the insight that the participants varied in their sensitivity to the room, leading to the incorporation of the second factor in which private body consciousness was measured and leading to the study’s conclusions. The end result is a richer theoretical understanding of the idea that disgust can cause harsher moral judgments but that this effect will likely vary across people at least by differences in what causes them to experience disgust.

In considering this type of design, it is important to remember that when non-manipulated independent variables are participant variables, they are by definition between-participants factors. These variables are generally assumed to be static, which is why they are measured instead of manipulated (unlike mood, for example). As long as one independent variable is manipulated, the design is still considered an experimental design overall, no matter how many other non-manipulated factors are included. However, conclusions about the non-manipulated variables need to incorporate the fact that these were not manipulated. We would want to avoid statements such as “high private body consciousness caused harsh moral judgment in a messy room” because it implies a causal effect on a variable that was not controlled. We would prefer to state the conclusion as a “messy room caused harder moral judgments in participants with high private body consciousness.” As we will review in Chapter 16, non-experimental relationships among variables are more difficult to interpret due to needed to consider and attempt to rule out alternate explanations.

## Hypochondria and Memory for Health-related Words

Another example of a design with one manipulated factor and one non-manipulated participant variable is a study in which participants were exposed to several words that they were later asked to recall (Brown, Kosslyn, Delamater, Fama, Barsky, 1999). The manipulated independent variable was the type of word. Some were negative health-related words (e.g., tumor, coronary), and others were not health related (e.g., election, geometry). The non-manipulated independent variable was whether participants were high or low in hypochondriasis (excessive concern with ordinary bodily symptoms). The result of this study was that the participants high in hypochondriasis were better than those low in hypochondriasis at recalling the health-related words, but they were no better at recalling the non-health-related words.

## Non-Experimental Studies With Factorial Designs

Thus far we have seen that factorial experiments can include manipulated independent variables or a combination of manipulated and non-manipulated independent variables. But factorial designs can also include only non-manipulated independent variables, in which case they are no longer experiments but are instead non-experimental in nature. Consider a hypothetical study in which a researcher simply measures both the moods and the self-esteem of several participants—categorizing them as having either a positive or negative mood and as being either high or low in self-esteem—along with their willingness to have unprotected sexual intercourse. This can be conceptualized as a 2 × 2 factorial design with mood (positive vs. negative) and self-esteem (high vs. low) as non-manipulated between-participants factors. Willingness to have unprotected sex is the dependent variable. But because neither independent variable in this example was manipulated, it is a non-experimental study rather than an experiment. This is important because, as always, one must be cautious about inferring causality from non-experimental studies because of the directionality and third-variable problems. For example, an effect of participants’ moods on their willingness to have unprotected sex might be caused by any other variable that happens to be correlated with their moods.

## Hypotheses in Factorial Designs

The primary goal of using a factorial design is to look for **interactions** among the design factors. An interaction is defined as one of the design factors modifying the effect of another design factor. For example, in the very first experiment diagrammed above, the effect of a cell phone on driving quality might be moderate during the daytime, but much larger at night. We would then say that the time of day influences the effect of the cell phone. Factorial designs are always designed to explore the interaction of factors. If we simply wanted to look at the effect of cell phones on driving and time of day on driving independently, we would run two parallel studies that each had a simpler 2-condition design.

At the same time, in the evaluation of the results of a factorial design, we have to systematically consider all the embedded hypotheses. In a 2 x 2 design, there are three hypotheses that are automatically being tested. We describe these as two **main effects** and the **interaction term**. In this example, one main effect is the overall effect of cell phone use on driving, but note that this is evaluated while ignoring any effect of time of day. Main effects measure the overall impact of that factor’s levels on the DV independently of everything else (the other factor or any interactions). A second main effect in this design is the effect of time of day on driving quality not including any effect of cell phone use. The technique for visualizing these main effects is to calculate **marginal means** from the results, which will be discussed in Chapter 10. Although the goal of the experiment may be to examine the interaction between factors, the results should always be presented comprehensively and include the main effects and interactions of interest.

The number of embedded hypotheses goes up quickly as design complexity is increased. For a 2 x 2 x 2 design, we now have 3 main effects and 4 interaction terms to consider. The main effects are one for each of the three factors. However, we now have potential interactions between the first and second factor, the first and third factor and the second and third factor. Then there is a potential three-way interaction among all the factors. In Chapter 10, we will review how to interpret results from factorial designs, identify the most common kinds of interactions and how to connect these to the experimental hypotheses.

## Key Takeaways

* Researchers often include multiple independent variables in their experiments. The most common approach is the factorial design, in which each level of one independent variable is combined with each level of the others to create all possible conditions.
* Each independent variable can be manipulated between-participants or within-participants.
* Non-manipulated independent variables (gender) can be included in factorial designs, however, they limit the causal conclusions that can be made about the effects of the non-manipulated variable on the dependent variable.
* In a factorial design, the main effect of an independent variable is its overall effect averaged across all other independent variables. There is one main effect for each independent variable.
* There is an interaction between two independent variables when the effect of one depends on the level of the other. Some of the most interesting research questions and results in psychology are specifically about interactions.
* A simple effects analysis provides a means for researchers to break down interactions by examining the effect of each independent variable at each level of the other independent variable.

## Exercises

* Practice: Return to the five article titles presented at the beginning of this section. For each one, identify the independent variables and the dependent variable.
* Practice: Create a factorial design table for an experiment on the effects of room temperature and noise level on performance on the MCAT. Be sure to indicate whether each independent variable will be manipulated between-participants or within-participants and explain why.

Brown, H. D., Kosslyn, S. M., Delamater, B., Fama, A., Barsky, A. J. (1999). Perceptual and memory biases for health-related information in hypochondriacal individuals, 67–78.

MacDonald, T. K., Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self-esteem lead to risky health behaviors? , 299–306.

Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological science*, *19*(12), 1219-1222.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and social psychology bulletin*, *34*(8), 1096-1109.