3 Experimental Control

We have described the process of setting up an experimental design as starting with the high-level constructs and then implementing operational definitions that allow us to create an experimental procedure that assesses the effect of an independent variable on a dependent variable. The previous chapter discussed some aspects of creating a measured operational definition that can be used as the dependent variable. In this chapters, we will consider issues and methods for creating effective independent variables that will allow us to draw strong conclusions from our research studies.

The goal of experimental design is to be able to draw strong, valid conclusions from the results of our studies. However, to illustrate the first step in design, we can take the perspective of reading a scientific report that we are led to through a social media link with the tag line

Listening to music improves scores on stressful classroom tests

The conclusion here is a causal statement, listening to music causes better test performance. That is an interesting and possibly useful statement and should make us immediately very curious about the design and the operational definitions used. We will want to know what the operational definition of the stressful classroom test is, but also what was used for music. One of the first questions we should ask is: **Compared to what?**

The conclusion statement attributes an effect to music, but as is often the case when we have only the conclusion, we do not know what music is being compared to. To really understand the result, we will likely have to find the original scientific report, read the procedure section and see how they describe their **control condition**. In this case, the control condition would result from the operational definition of the independent variable. We would expect to have a description of the condition in which participants were exposed to music and in that section also a description of the condition in which there was no music.

Returning to the perspective of designing a study, we might start with a similar hypothesis such as: Music helps me study. Notice that this is also a causal statement that just like the tag line, leaves out a lot of information about the independent and dependent variable. However, since we are thinking of designing a study to test our hypothesis, we have to come up with the operational definitions for these variables. Doing it well means that if we obtain a reliable effect in data we collect in our study, we will be able to confidently assert that our hypothesis was correct. In this Chapter we will discuss principles of design that aim to guide us to doing it well.

The Mozart Effect

Studies of the effect of music on cognition are often referred to as related to a hypothesized Mozart effect. This has essentially become a colloquial term for an old idea that classical music has some benefits for general cognitive processes. That specific idea has not replicated effectively but more modern studies have found small but reliable effects of ambient music on specifically spatial cognition tasks and may also be helpfully calming in some stressful environments.

First we will have to establish

what we mean by *music* and also what we will use as the *non-music* contrast. *Music* could refer to a specific piece of classical music, or perhaps popular music preferred by the participants. The control condition could be silence, or could be white noise, or soothing natural sounds. There is not one right answer to how to pick a correct operational definition here. Most of the alternatives are different, equally interesting studies.

One we select the music and non-music conditions, we will be planning to collect data from a group of participants. This is when we need to consider the question of experimental control. When we do our data collection is there anything else going on around the participants that might affect how we administer the conditions or measure our DV? Is there other noise around? Might we get interrupted? How consistent and reliable is the technique for playing the music?

Anything not part of our planned design that affects our study is referred to as an **extraneous variable**. Experimental control is fundamentally about identifying as many extraneous variables as possible and designing our study to minimize the effects of these.

These uncontrolled extraneous variables can affect the quality of our research process in two main ways. Random external influences that interfere with carrying out the research procedure will tend to produce **reliability** problems. Technically, these increase the variance in the average scores on our dependent variable which makes it more difficult to obtain statistical confidence that our IV affects the DV. For example, if there are interruptions occurring randomly for some of the participants in the study, their DV scores might be higher or lower because of the interruptions rather than our control over the music being played.

That is a problem for our research, but an even greater problem occurs when there is an extraneous variable that systematically varies with our IV. This creates an **experimental confound**, which is a serious problem with the validity of any conclusions we would like to draw from our research.

Confounds are going to be the most important problem we are concerned with when evaluating the validity of an experiment because they essentially ruin the ability to get a confident conclusion. As an example from our hypothetical music design, if we used different audio volume for our music and white noise stimuli with the white noise being painfully loud, we would see an effect on studying but it would not be a positive effect of music, it would be a negative effect of loud distracting white noise. Technically, the audio volume here is an extraneous variable and since it varies with the IV by always being louder for the white noise, it has confounded our experiment.

More commonly, extraneous variables are related to things like the testing context, whether there are distractions around, details of the technology or even things like time of day. These aspects can affect our data collection but if they are random and occur for both conditions roughly equally, they do not confound our study. In the laboratory environment, where we have reasonable control over the environment, there are two main techniques to managing these extraneous variables: **constancy** and **counter-balancing**.

In this chapter, we will introduce the basics of setting up our IV in experimental design with consideration of all the extraneous variables we can identify. We will review the particular danger of confounds and the impact of these variables on the reliability and/or validity of experimental design. In Chapter 4, we will extend this discussion to the specific challenges of things that differ across the participants in our study and the basic approach to establishing rigor through a well-planned experimental procedure.

Learning Objectives

1. Understanding the independent variable in experimental design
2. Define what a control condition is, explain its purpose in research on treatment effectiveness, and describe some alternative types of control conditions.
3. How to construct two treatment conditions as the IV for a study and how this might be extended to more complex designs
4. Extraneous variables: factors that affect the DV that were not part of the experimental design
5. Managing extraneous variables via constancy: keeping things as consistent as possible across levels of the independent variable
6. Managing extraneous variables via counterbalancing: if factors cannot be kept constant, distribute them evenly across the independent variable
7. Confounds: variables that reduce the internal validity of an experiment

Treatment and Control Conditions

For many people, their introduction to psychological research is by seeing findings reported from a study designed around having a **treatment** condition and a **control condition**. These kinds of studies are essentially **intervention** studies where something is done to improve things or fix a problem. This kind of study is very common in health and medical research as well and we will use those domains for examples of design even though the main application will be studies of psychological constructs. We will return in Chapter 17 to questions specific to intervention research, particularly implementation and ethical issues.

The music example above can be seen as a kind of intervention study as a test of the hypothesis that *music helps with studying*. Music is the intervention in question and improved studying would be the benefit. The use of the example is to focus our attention on the control condition that must be present in the design in order for there to be reliable scientific evidence for the statement. In our above example, we considered using white noise as the comparison to music. As a result, our more accurate statement of the results of testing the hypothesis would be *music helps with studying compared to white noise*.

The same style of thinking should be applied to other hypotheses like *daily meditation helps with focus in the classroom* or *exercise helps to sleep better*. Those statements also have hidden control conditions to which an intervention was compared. When reading research, we look in the sections of the scientific report where the procedure is described to find the operational definition of the control condition. When designing research, we consider the inference we are hoping to draw and the best comparison condition to establish our conclusion.

The rest of the basic design for this kind of study is very simple. We administer the treatment to one group of participants, usually half of the total group. We administer the control condition to everybody else. Then we assess the dependent variable and see if the treatment affected those scores.

Of course, this approach being simple and effective depends on solving the measurement problems that were the main topic of Chapter 2 as well as

_______ with this one simple trick!

In the public sphere, interactions with findings from science often come packaged as statements like this where the blank could be filled in with lose weight, make more friends, or get better grades. If there is any science behind the statement it will be based on research that uses an intervention design. If it is good science, there will be an appropriate control condition, good sample size, no confounding variables, statistical reliability and a consistent rigorous study procedure. If not, then it is just click bait.

the experimental control questions which are the focus here of Chapter 3.

Before we generalize from treatment and control to more general designs with two levels of an independent variable, it is worth noting one specific design element associated with these kinds of designs, the **placebo** effect. Intervention research often needs to explicitly discuss the possibility of the control condition inadvertently affecting the dependent variable through the expectations of the participants in the research study. We will return to this idea in Chapter 4 under the general consideration of **demand characteristics** embedded in experimental design and standard approaches to controlling these.

Independent Variable with Two Levels

The basic framework we are using to examine research methodology in psychological science is to quantify how manipulation of an **independent variable** affects measurement of a **dependent variable**. Using treatment studies as an example, we see that at a minimum, our independent variable needs to have two conditions which we will more generally refer to as **levels** of the variable.

Our Experiment 1 illustrates the more general approach. To assess how deeper semantic processing of words leads to better memory for those words, two levels of depth of processing are compared. The deep condition, where words were rated by the participants as how much they liked them, is designed to have more depth than the comparison shallow encoding condition of counting vowels. Rather than thinking of the design as based on a treatment and control, we describe the independent variable as being an operational definition of depth of processing and use a procedure that has two different levels of depth.

Choosing an effective control condition is not always a straightforward process as the earlier music example showed. Research in behavioral health measures aimed at interventions such as physical activity to improve health outcomes is an area famous for the difficulty in constructing control conditions. As a simple example, a straightforward intervention is to ask participants to add exercise to their daily routine to improve cardiovascular health. When the control condition is simply not adding exercise, it is important to also understand what activities are potentially being replaced by exercise. If the control group is not just sedentary but also engaged in unhealthy eating behaviors that might have been replaced by exercise, the intervention may not work in the manner hypothesized. That is, it may improve health by reducing unhealthy eating instead of via a direct effect of exercise on the body. Since we usually want to understand why our study worked, we might prefer using two levels of the independent variable instead. Here that might be comparing a high level of exercise, like running for 30 minutes, with a lower level of exercise, like stretching for 30 minutes.

To be clear, the use of terms like treatment or levels of the independent variable is merely a matter of terminology. There is no conceptual or design difference between descriptions that prefer one set of terms versus the other. There is even an additional synonym for the manipulated variables of an experiment where these are referred to as **factors** in design. Which of these terms gets used reflects customs in different sub-areas of science. The term factors is typically used in descriptions of more complex designs than we will start with in this chapter. When there are multiple independent variables being manipulated by the experimenter, these are described as **factorial design**, a topic discussed in great depth in Chapters 10 & 12.

When more than two levels of an single independent variable are contrasted, this design may also be called a factorial design. An independent variable with three levels adds only a little complexity to questions of experimental control. However, it adds substantially to the problem of drawing inferences from the data, partly due to needing more complex tools for statistical analysis. In addition, three levels implies at least three key comparisons that will need to be evaluated and interpreted. If there are conditions A, B and C in a design, we have to contrast A versus B, B versus C and A versus C, each with their own statistical contrast and conclusions to be drawn about reliable differences found. Our discussions of design will focus on just two levels of the independent variable to keep this aspect of design and interpretation simple to start.

For the basics of experimental control, we will also temporarily hold aside the additional complexity brought on by designs in which participants each experience all levels of the independent variable. These designs, called **within-participant**, can also have an independent variable with two levels but all participants get both levels. Experiment 1 could have been administered this way by asking participants to rate some words for liking and count vowels for other words. In Chapter 7, we will discuss the many strengths of this approach to research and techniques for managing the additional challenges created by considering the possibility of order effects in administering the independent variable.

If everything else about experimental control goes as planned, differences in the scores on the dependent variable can only have arisen from the different experience of the participants across the levels of the independent variable. That is what allows us to make causal statements that the different levels of the independent variable caused different scores on the dependent variable. However getting everything to go as planned is not as simple as it looks and requires a research procedure designed with the necessary experimental control.

Extraneous variables

Virtually everything we need to worry about with respect to planning an effective experimental design aimed to produce reliable outcomes and valid conclusions boils down to identifying and managing **extraneous variables**.

Extraneous variables reflect anything going on around or during the experiment that could affect scores on the dependent variable that were not part of the experimenter's design. In most simple experimental designs, we will be planning to collect data on samples of participants roughly including

20 to 60 people. The conditions under which these participants are in the experiment will vary across those data collection points. They could vary in time of day, location, distractions around them, etc. If participants are completing an experiment through an online system, they could be interacting with the system from a variety of devices, different internet connectivity strengths, different interfaces like touchscreen or keyboard.

We will rely heavily on intuition and our knowledge of psychology as people to identify as many relevant extraneous variable as possible for the designs we consider. As you develop specific knowledge of research within a specific area of psychology, you will learn about the variables that commonly have to be considered in that domain. It can be tricky to find these in research areas that we are less familiar with. Since this class is about general design, we will tend to not focus in detail on domain-specific details. As an example, unless you are familiar with memory research with word lists, you are likely unaware that characteristics of words themselves influence memory for those words. Words that are very uncommon, termed *low frequency*, tend to be much more memorable for designs like the one we used in Experiment 1. We will see that the specific words in that study were selected with awareness of this extraneous variable and designed to minimize the potential impact of this aspect of the stimuli.

There are always a large number of possible extraneous variables implicit in any experimental design. Our general approach to evaluating designs will be based on reviewing the described procedure and then brainstorming as many of these as we can come up with. Then evaluate which are likely to be affecting measures in the study, likely discarding most of the candidates we consider. If we identify a plausible variable that seems likely to have a substantial impact on the dependent variable measure, we will go back to the described procedure and verify that this element was handled effectively. If not, there may be a problem with the conclusions drawn from the study.

We use the same process for developing a new experimental design except that as experimenters, we are responsible for putting together the procedure for the planned design and handling these variables. We will discuss two basic approaches to keeping these variables from weakening our research methodology here. It may be worth mentioning here that a very useful approach to the problem of identifying all the possible extraneous variables in a proposed research design is to follow many of the procedure elements in a published scientific study. Successful studies have generally demonstrated how to avoid the problems of extraneous variables. Later, when we discuss creating research proposals for potential class projects, the recommended technique will be to take an existing, working design and add one new thing to it to help increase the probability of a successful result.

In this chapter we are focusing on extraneous variables associated with the environment, procedure and stimuli. There are also clearly a lot of aspects related to the participants in the study that can affect the dependent variable. For our Experiment 1, we might hypothesize that some participants are better at memory for words, some were better able to pay attention during study, some were stressed or had not slept well the night before. These **participant variables**, which often reflect **individual differences** between people, are managed slightly differently and will be discussed in Chapter 4. Conceptually, these concerns influence the accuracy of experimental design in the same manner as other extraneous variables.

We will consider the impact of these extraneous variables as they affect the **reliability** and **validity** of our experimental design. Note that these terms here are being used in their formal, technical sense. Reliability specifically refers the statistical evaluation of experimental results. That is, the data indicate that we have sufficient confidence the results could not have happened by chance. This is equivalent to saying we believe the result will **replicate** if the experiment were carried out again. Validity specifically refers to whether the inference is accurate that the independent variable affected the dependent variable. It might seem like these two attributes would tend to go together, but they do not always. We can have an experimental result that is reliable, but not valid. And likewise, a valid hypothesis for which the data is not reliable in a study.

Errors arising from design mistakes

One set of terminology used in describing experimental errors is to describe errors as either **Type 1** or **Type 2**. It is more important to understand the implications of these errors and how to spot them than be specifically familiar with the terminology, although the short hand terms are frequently helpful.

A **Type 1** error is a **false positive** claim where we believe the independent variable has causally influenced the dependent variable, but the claim is wrong. This happens when there is a failure of **validity**, often also described as an error of **internal validity**. This error is a serious problem for science since it asserts an incorrect claim, which if believed can cause damage when people rely on the claim to influence their behavior subsequently. Much of what we do to establish a rigorous experimental design is aimed to minimize the possibility of this kind of error. In consideration of extraneous variables, the key problem we aim to avoid is having a **confound** in our study, a **confounded variable** with our independent variable (see below).

A **Type 2** error is a **false negative** in which our study does not reliably support a claim about the independent variable affecting the dependent variable. This generally reflects a failure in **reliability**, which we will interpret mainly on statistical grounds. This happens when an experiment does not appear to work, that is, across the two levels of the independent variable, the differences in the measure of the dependent variable were not large enough support a claim of a finding. This is an error when it turns out later that the independent variable normally does affect the dependent variable but the particular study carried out did not observe the typical effect. That can happen simply due to poor luck related to happening to observe higher levels of variability in performance than usual. It can also happen when there are a lot of uncontrolled extraneous variables that were not managed properly in designing the experiment.

In general, we prefer a Type 2 error to a Type 1 error since a lack of statistical reliability will often lead to results not being published and therefore there is no danger of people relying on an inaccurate scientific finding. Note that this is different from concluding that the independent variable never affects the dependent variable, which would be a true negative. As we will see in Chapter 5, standard statistical models for scientific inference are not well designed to test this kind of null hypothesis. Difficulty in establishing no effect is part of why a Type 1 error is difficult to correct, as it requires correcting a finding with a null finding. Alternate statistical approaches have been proposed to extend our inference models, but none have been broadly adopted and these are outside the scope of our methodological discussions here.

Most of the time, extraneous variables do not affect the internal validity of a study. They create noise in measures that can lead to a failure to reject the null hypothesis statistically so that the experiment does not produce a result supporting a conclusion. When that happens, you do not know initially if you have experienced a Type 2 error, or simply that your hypothesis is wrong. The only thing you know for sure is that your experiment did not work. However, a critically important aspect of experimental design is to be as careful as possible to avoid the possibility of an experimental confound in your design. These cam lead to the much more problematic Type 1 error, a false claim.

Confounds

For an extraneous variable to be a confound, it has to vary with the independent variable. That is, the extraneous variable changes so that it exactly matches the independent variable. When this happens, we can no longer be confident that the intended independent variable was the cause of any changes we observe in the dependent variable. If a reliable effect is observed, it could have been caused by the extraneous variable. It is also possible that it was caused by the independent variable, but once there is a confound, there is no way to know what caused the effect.

We can illustrate the problem with our music example above. First we select operational definitions for music and the comparison, non-music, condition, which we suggested could be white noise. Implementing these, we would plan to play audio clips and realize we had to set the volume these clips are played at. The volume here is a potential extraneous variable as it will affect perceptions of the intended independent variable and then influence the dependent variable scores. The worst thing you could do would be to have the volume for the music and the white noise be very different. If the white noise was played at a loud, obnoxious volume and the music at a pleasant, moderate volume, the extraneous variable volume would be confounded with music. If the scores on the dependent variable differed between conditions, we would have no way to know if it was the music or the loudness that caused the effect. Our experiment is confounded and we would describe any attempt to state an effect of music to have a problem with internal validity. If we attempted to claim from these data that music improved studying, we would be at risk of making a false claim, a Type 1 error.

One of the first elements of your design that need to be considered for potentially confounding extraneous variables are the stimuli to be used in the experiment. If we are using pictures in our study, we want all the pictures to be as similar as possible across conditions. If we ask participants to read stories about altruistic behavior, we want to be sure everything else about the stories they read is as similar as possible. Ideally, the only thing that differs is the exact variable we want to use as our independent variable.

We can also consider the broader contextual elements of our data collection. If there are multiple research team members engaged in data collection, they should not each run different conditions. If you are collecting at different times of day, avoid collection one condition in the morning and the other in the afternoon.

If it seems like it should be easy to avoid these kinds of problems, most of the time it is! Below we will provide two terms to describe very simple and straightforward ways to plan for these kinds of extraneous variables. It is necessary to go through the process of identifying the potentially confounding variables and controlling them since not doing so is essentially catastrophic for being able to draw any conclusions from your research study.

Internal Validity

The term **internal validity** is used to characterize an experimental design that will be able to test the underlying hypothesis. Any major problem that impairs the ability to draw a conclusion from the experimental data is a problem with the internal validity of the study. In addition to confounded variables, one way this can happen is if there is a mistake in the operational definitions. If they do not accurately reflect the underlying construct, the main inference about the constructs cannot be drawn from the data. Internal validity challenges are closely related to the problem of Type 1 errors.

This issue is distinct from **external validity**, which reflects the degree to which the conclusions can be applied to participants outside the research lab, in the real world. External validity generally depends on the methods of sampling participants, that is, how they are found and recruited into the study. This issue will be discussed in depth in Chapter 13, but as a preview, you can consider the concern being raised about the general dependence of psychological research on behavior measured from undergraduate students at major American universities. The question is whether the results obtained from university participants correctly predict the behavior of the broader population and whether we need to consider broader sampling or limiting the expected breadth of our conclusions.

If you suspect there might be a problem with external validity, that does not mean there is a problem of internal validity. For example, in our Experiment 1, the participants are university students. This not a problem with internal validity, nor a confound in the design. There might be a **limitation** to our conclusions that would cause us to raise the question of whether better memory from deep encoding only occurs with this specific type of population. In Chapter 4, we will consider when characteristics of the participants can actually affect internal validity but this is a separate issue from external validity.

Non-confounding Extraneous Variables

Most of the time, the many variables that vary across the course of collecting data for an experiment are not confounded with the independent variable. Things like time of day, or time of year, the weather, the details of the testing room, the social skills of the researchers supervising the data collection. If these differences are happening at random, they will affect the dependent variable, but will occur roughly equally often across both levels of the independent variable. Mainly what this does is increase the observed variance in the data, posing a challenge to the reliability of our results by making it more difficult to observe a robust difference in the dependent variable scores.

Any measure derived from human participants is going to have variance in performance associated with it. This is embedded in our statistical model for determining reliable effects of the independent variable. The difference in the average scores for the participants in each condition must be sufficiently large compared to the variance for us to have confidence it did not happen by chance.

Conceptually, variance in measured scores results in part from measurement error, which reflects the important idea that no quantitative operational definition is ever perfect. Another important component of this variance is the random noise caused by these extraneous variables. If there are many of these and they have large impacts on the measurement of the dependent variable, then extraneous variables can create **reliability** problems for an experiment design.

If the extra variance leads us to be unable to conclude that there was an effect of the independent variable on the dependent variable, then we may have run into a **Type 2 error.** It appears our experiment did not work, but we do not know for sure if our hypothesis was actually false. As noted above, this kind of error is less costly than a false positive created by a confound, but it still means the time and effort put into carrying out the study was not put to best use.

Control of Extraneous Variables

The principles for implementing best practices for reducing the effect of extraneous variables are simple in theory. Once the variables have been identified, keep as many as possible constant across conditions following the principle of **constancy**. Anything that cannot be kept constant but can be controlled, **counterbalance** across conditions so that it occurs equally often across levels of the independent variable. These two basic techniques remove the possibility of extraneous variables being confounds and maintain the internal validity of the study.

Constancy

As much as possible in any experimental design, keep things constant across the levels of the independent variable. This is the preferred technique for extraneous variable as it provides the best opportunity to observe a reliable effect of the independent variable. If the only aspect of the study that differs across conditions is the independent variable, then we can be very confident that changes in the dependent variable were caused by the manipulation.

Looking at the structure of Experiment 1 in Chapter 1, it should be clear how much of the presentation of that experiment was designed with constancy in mind. Everything possible about the overall look and feel of the interaction with the word lists was kept the same, except for the instructions about how to interact with the words. Maybe it would not have mattered if the word font, or font size differed across conditions, but if the readability of the words affected memory, then we would have had a major confound in the design.

It can be surprising to students who get to participant in psychological science research how meticulously detailed data collection procedures typically are. Many studies have carefully written scripts describing how the research team interacts with participants through the data collection process. This is done to keep interactions as constant as possible across conditions as well as to manage **demand characteristics** of psychological research, which will be discussed in Chapter 4.

The importance of this level of experimental control is also seen when research is reported through. In Chapter 6, we will see that the Methods section of an APA-formatted research report includes a lot of this meticulous detail so that the reader can evaluate whether sufficient experimental control was used to justify the conclusions of the study.

The ability to impose a high level of control is a hallmark of laboratory experimental research. Studies done in well-controlled conditions will have the highest level of internal validity. Later, we will consider and contrast approaches used in field research where it is necessary to give up a lot of this control in order to increase our external validity and confidence that the findings apply in situations outside the laboratory.

Counterbalancing

For any factors that cannot be kept constant, distribute how these are implemented equally across conditions. For example, if participants are being run throughout the day, collect data from both of the experimental conditions equally early and late in the day to avoid confounds due to circadian (time of day) effects. If it is necessary to have multiple experimenters, make sure they each contribute to data collection in each condition. If the stimuli are presented in different orders to participants, make sure the orders are used equally across the conditions of the study.

Counterbalancing is focused on making sure the extraneous variables do not confound the study. It allows for the potential that these variables may contribute to measurement noise and increased variance in the dependent variable. When the full control of constancy cannot be achieved, we always prefer the risk of a Type 2 error, where our experiment does not achieve a reliable result, to the risk of a Type 1 error, where we draw an incorrect conclusion due to a confound.

Practically, implementing a counterbalancing procedure can be as simple as alternating the administration of experimental conditions. In Experiment 1, students may be completing this study at the beginning of a class session. Early arrivers to class will have a slightly different experience than late arrivers. Students who arrive to class slightly later will potentially be completing the study under a sense of time pressure and aware that the rest of the class is waiting for them. Since we cannot control when students arrive to class, the experiment is implemented to alternate experimental conditions for each person who starts the study. If one student receives the *deep encoding* instructions, the next student receives *shallow* instructions. This distributes the variable of arrival time across the experimental conditions and removes the possibility of this being confounded with the independent variable.

In some cases, we may not be able to control variables enough even to counterbalance them carefully. In Experiment 1, participants may have been completing the study on a laptop, or on their personal phone which often have very different size screens. The online administration of the study was not done in a manner that allowed control of this variable so instead, we relied on the technique of **random assignment** to distribute this variable across our conditions.

There will always be a number of variables that are outside our control that will end up randomly in one condition or another. An important class of these variables in psychological science is differences among our participants. For Experiment 1, some people might be better at memory for word lists than others. In Chapter 4, we will discuss these kinds of **participant variables** and see that random assignment is the primary technique for experimental control. This approach does not reduce the internal validity of our experiment. In fact, our statistical procedures are designed around exactly the idea that the observed difference between conditions is larger than the difference that would happen if it was solely due to random assignment of these extraneous variables across conditions.

Summary

When planning a research study, or reading about a completed study, the standard method to try to identify potential confounds is to try to think of as many extraneous variables as possible that might affect the dependent variable. There will generally be quite a few, but most or all of these will not vary with the independent variable so we do not have to worry about them reducing the internal validity of the study by creating a confound.

Figuring out all the relevant extraneous variables can be challenging and benefits from knowledge and experience with related research findings. Practically speaking, there are generally a set of variables related to the stimuli used in the experiment and testing conditions that can be managed in order to both avoid confounds and minimize noise in the measure of the dependent variable.

One of the challenges to psychological research is that we can never be sure we have found all the possible extraneous variables. In fact, one major avenue of scientific discovery is finding new factors that affect our dependent variable that was not previously expected. Psychological science generally advances as a series of studies that build on each other using slightly different operational definitions and methods of experimental control. This both builds confidence in the main conclusions drawn about the underlying constructs and allows for gradual identification of other factors that affect psychological processes.

Key Takeaways

- An **extraneous variable** is any variable other than the independent and dependent variables.
- A **confound** is an extraneous variable that varies systematically with the **independent variable**.

- Studies are high in **internal validity** to the extent that the way they are conducted supports the conclusion that the independent variable caused any observed differences in the dependent variable.
- Experimental methods are high in internal validity when confounds are avoided because the manipulation of the independent variable lets us infer that is what caused the observed difference in the dependent variable.
- A **Type 1 error** is a false positive claim where a researcher mistakenly thinks the independent variable reliably influences the dependent variable but it does not.
- A **Type 2 error** is a false negative claim where a researcher thinks the independent variable does not affect the dependent variable but it does in truth.
- Extraneous variables that do not confound the study increase variance in observed performance increasing the probability of a Type 2 error.
- **Constancy** is a method for reducing the effect of extraneous variables by as much as possible keeping everything the same across levels of the independent variable
- **Counterbalancing** is the approach to use for anything that cannot be kept constant. Distribute extraneous variable conditions across the levels of the independent variable to keep this element from being confounded with the independent variable and reducing validity of the experiment.

Exercises

Question 1: Laughter is the best medicine

Imagine you have just read an article in the newspaper describing a scientific study in which researchers found that people who laugh a lot tend to have lower blood pressure, stronger immune systems, feel less stressed out.

Considering the problem of extraneous variables and potential confounds, give an alternate hypothesis for how this relationship might be observed without supporting the authors' conclusion. Note that this requires a statement consistent with the data, not consistent with the conclusion.

Outline an experimental approach to this question that would more directly test the hypothesis. Provide an example of an operational definition of the IV, the DV and what you would expect to find if laughter positively affects health.

Question 2: Briefly answer the following questions about experimental control from our Experiment 1:

- Why have both groups read the same words?
- Why have 1-5 scales for responding for both conditions?
- Why require the word to be on screen for minimum 3 s?
- Does it matter if the trivia questions use words from the study list?

4 Experimental Procedure

In each of the examples given at the beginning of the first three chapters, the process of trying to identify all the extraneous variables should encounter the idea that people themselves differ importantly on each of the various measures. Not everybody is the same when it comes to memory for words, self-esteem, or reactions to music. Individual differences on these aspects of psychology are considered **participant variables** for the purpose of experimental design. Classic examples to consider are the age and gender of the participants, which may affect some psychological constructs.

It is important to acknowledge that our participants will differ but it would be impossible to try to identify every possibly way they differ in order to apply our experimental control tools of constancy and counter-balancing to these variables. There is a large area within psychological science, **personality research**, aimed at identifying, understanding and characterizing aspects of how humans differ from each other. Psychological science has a core philosophical tension between the attempt to draw broad conclusions that we believe are true for all humans while also acknowledging that many of these statements will apply differently to some individuals.

Since we cannot control participant variables, we mainly aim to keep them from being confounded with the independent variable. The technique for this is to **randomly assign** participants to conditions. This implements counterbalancing as long as we do not somehow get unlucky that participants similar on some variable all end up in the same condition. Making this kind of unlucky circumstance unlikely is one of the main reasons we carry out psychological research on sizable groups of participants rather than a few individuals. Mathematically, once we have groups in the range of 30 participants or so, it is exceedingly unlikely that random assignment would lead to unbalanced participant variables. Such a low probability event is even less likely when a study gets **replicated** a second time with a new group of participants.

In recent years, there has been some concern in psychological science about important findings being difficult to replicate with new studies. The fear is that some claims are actually Type 1 errors and need to be discarded and underlying theoretical ideas revised. However, it is difficult to be sure of this as a failure to replicate might itself be a Type 2 error reflecting some aspect of weak experimental control in the attempted replication. However, the most interesting cases of failures to replicate is when they reveal new variables that affect the constructs and these often come from participant variables.

As an example, you may be familiar with the idea of stereotype threat in which exposure to a group-based statement of cultural expectations of poor performance can actually create poorer performance in that group. Stereotype threat research can be carried out with a simple design with two levels of an independent variable. In one condition, participants are exposed to the stereotype threat content and in the other condition, participants are exposed to control content that does not mention the stereotype. The dependent variable is measured performance on a related test. When a group participants score lower on the test, we can infer that exposure to the stereotype caused lower scores. Note that these results do not reinforce the stereotype because the group of participants not exposed are not affected, even though they are from the same group.

However, not every study of stereotype threat has been found to produce a reliable effect, leading to questions about the robustness of the phenomena. In Aronson et al. (1999) some insight into this variability was provided in a study that examined stereotype threat on math performance but further asked participants how important math was to them, a participant variable they defined as *math-identification*. For students who self-reported that math was extremely important to their identity, stereotype threat was found to impair performance. Their study was also notable in that they used threat stimulus applied to white males who were exposed to the stereotype that *Asians perform better on math tests*, which additionally showed the influence of stereotype threat on a group normally seen as privileged. Students who reported that math was important to them were negatively affected by the stereotype, replicating the main hypothesis. However, students who reported low math-identification did not score lower on the test, indicating that they were not affected by exposure to the stereotype.

This finding implied that our understanding of stereotype threat needs to incorporate the idea of individual identity and how these interact. This insight likely emerged from researchers puzzling over a failure to replicate the simpler study and developing a theory that there was an unconsidered participant variable involved. Phenomena like stereotype threat are an important part of understanding how inequality can be inadvertently embedded in an educational context. It is therefore important that when the phenomena is found to vary across experiments that it not be discarded, but explored further for better understanding. Additional research can then build on these findings to test and understand ways to combat harmful effects of stereotypes on students in the classroom.

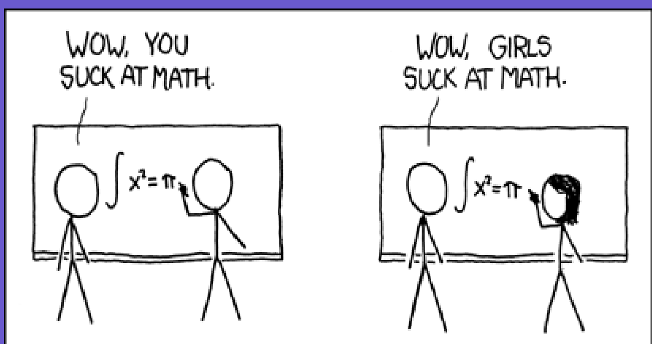
This example illustrates a difficult aspect of coping with extraneous variables in experimental design. The important factors are often not known in advance of the research. It can take a lot of experience and expertise in the specific research domain to learn where design problems might emerge from before new research can be done to extend the known theory about the main constructs for the study.

Stereotypes and bias are also well-known to be challenging to study in laboratory conditions because of effects related to **demand characteristics**, which we will also discuss below. These have to do with participants modifying their behavior due to the awareness that they are participating in a research study. Participants knowingly aware they are part of a study of stereotype bias may monitor their behavior more carefully in order to avoid any accidental implicit shows of bias. That may pose a difficult challenge for constructing a research protocol high in internal validity.

Challenges even in simple experimental design reminds us that psychological science is not a list of facts to be memorized, although it may in some cases first be taught this way. It should be seen as a process of successive understanding as scientific studies build on each other to gradually improve our overall understanding of human behavior, cognition and emotion.

Bias in studying bias

Methodologies for studying stereotype threat are often excellent examples for simple experimental design, but teaching them seems to run the risk of actually causing the effects they describe. Stereotypes are examples of misunderstanding of correlation and causation. They may describe current relationships that could arise due to historical expectations, but lead to misleading conclusions about why. Observed relationships in the world are too frequently assumed to reflect inherent differences instead of effects caused by environmental factors.



Learning Objectives

1. **Participant variables** as extraneous variables that may affect the dependent variable, validity and reliability of psychological research.
2. **Random assignment** of participants to conditions to best distribute differences on participant variables.
3. **Demand characteristics** in experimental research and the effect of awareness by participants of being in a research study.
4. Using **single-blind** methodologies to limit some kinds of demand characteristics.
5. **Placebo effects** as an example of demand characteristics in health and intervention research
6. Avoiding bias in scoring the dependent variable through **double-blind** methodologies
7. Implementing the experimental procedure using care and consistency in following the protocol for best experimental control

Participant variables

People differ in ways that might affect their behavior in an experimental situation. They may vary in aspects that are measured on the dependent variable. They may react or interact with the levels of the independent variable differently. These differences act like the extraneous variables discussed in Chapter 3. If the participant variables are confounded with the independent variable, this causes a threat to the internal validity of an experiment. If they are randomly distributed across conditions, validity will not be challenged, but these differences will contribute to variance in the quantitative measures in the experiment and may challenge reliability.

We have to treat participant variables slightly differently than some other kinds of extraneous variables because we cannot control or change these aspects of our participants. We have a lot of control over the stimuli in an experiment, the environment around the testing procedure, the manner in which the research team interacts with the research participants. For participant variables, we have one very effective technique: **random assignment** to conditions.

Random Assignment to Conditions

As long as participants are randomly assigned to conditions, individual differences should never confound the final result. It is tempting to worry that it is possible to get unlucky in our randomization and assign all the participants who are better at the task to the same condition. However, this is exactly what our statistical tools are designed to test. For all our statistical tools for deriving inferential statistics, the *p value* that we calculate is formally the probability that we accidentally observed the difference that occurred due to this random chance. The idea of accidentally seeing the difference is formally the same as the **null hypothesis** that there is no effect of the independent variable on the dependent variable. When we reject the null hypothesis, we explicitly consider and mathematically rule out the possibility that individual differences, or any other non-confounding extraneous variable accounted for our results.

It is important to note that for random assignment to work, it has to be carried out correctly and there needs to be an adequately large sample of participants recruited for the study. A good, simple rule-of-thumb is to try to have at least 30 participants in each of your experimental conditions, if possible. It isn't always possible to obtain that many volunteers, however, and 15-20 per condition also often works. Carefully estimating statistically adequate sample sizes is a process called **power analysis**, which can actually be a complex task. A full consideration of the factors that go into a power analysis is beyond the scope of this text, but we will touch on the key ideas in discussions of statistics in Chapter 5, sampling techniques in Chapter 11 and research proposals in Chapter 13.

Smaller sample sizes weaken the effectiveness of random assignment. In some specialized cases with restricted populations such as neuropsychological research or other kinds of case studies, it is not always possible to recruit large samples. In these cases, it may be necessary to use designs based on **matched participants**, where participant-based extraneous variables are assessed and explicitly balanced across the levels of the independent variable. This and related techniques were used in some older psychological science studies that pre-date the modern recommendations to use larger sample sizes. The challenge of matching procedures is the need to identify all possible participant-based extraneous variables and then have reliable measures of all of these prior to assigning conditions. It is generally much simpler just to randomly assign a large group of participants to conditions and trust that the statistical model will account for assignment luck.

Random assignment, properly carried out, will prevent individual differences from confounding an experiment. However, incorrectly following the randomization procedure can lead to embedding bias in a study. A very simple procedure for effective randomization is to alternate conditions for participants as they are recruited into the study. Later in this chapter we will discuss the importance for having a pre-planned randomization strategy, even a simple one, for avoiding any accidental bias in carrying out a research study.

Constancy in participant variables

In Chapter 3, we saw that our two main approaches to extraneous variables are counter-balancing and constancy. It is possible to apply something like constancy to participant variables by selectively recruiting participants keeping a variable constant. For example, many cognitive neuroscience studies of language limit participants to right-handed people, who generally have their language areas isolated in their left cerebral hemispheres. Left-handed people are more likely to have their language areas isolated in their right cerebral hemispheres or distributed across both hemispheres, which can change the way they process language and thereby add noise to the data. We might also select participants if our hypothesis is specifically about a subgroup, such as reactions to emotional stimuli among people high in anxiety.

Selective recruiting can increase the reliability of a planned study without causing a reduction in internal validity, but it lowers the **external validity** of the study—in particular, the extent to which the results can be generalized beyond the people actually studied. Typically this results in conclusions that are limited by including the recruiting criteria. Cognitive neuroscience studies of language typically have to include the statement *in right-handed participants* to reflect that selection of participants.

Historically, a great deal of early health-based research was done with insufficient attention to maintaining appropriate diversity in participant recruiting. In some studies, recruiting was entirely based on white males, leaving large gaps in the scientific literature about how these health interventions affected everybody else. The attempt to justify this at the time was that this reduced variability in participants, increasing the power to detect whether a health improving intervention was clinically effective. However, it should be clear that this also raises significant ethical concerns that these research studies were not being designed to provide benefit widely across the population. This is rare in modern research but provides a reminder that we would prefer our science to provide findings that apply as broadly as possible. So restricting participants on their characteristics is something we would generally prefer to avoid.

Concerns about how effectively even modern psychological science really capture the diversity of the human population have drawn attention to the fact that a lot of psychological research depends entirely on university students as participants. In the USA, Canada and Western Europe, it has been observed that the undergraduate population may not reflect the broader population in society. This issue has been described as an over reliance on **WEIRD** participant samples, which is an acronym for Western, Educated, Industrialized, Rich and Democratic. Asa result of this recruiting aspect, commonalities in social or cultural expectations in these participants may be implicitly embedded in many psychological research reports. The main implication of this idea is that there may be unknown extraneous variables that vary across social and cultural groups that affect behavior in ways we have yet to explore in research. That does not invalidate research that depends on WEIRD populations, but may affect applications of the findings to broader, more diverse populations.

One technique for increasing potential diversity of research is to use methodologies for collecting data online. Research on how online methodologies affect recruiting diversity is ongoing but suggests that these samples are at least more diverse across ages.

Practically, psychological science reflects restrictions in recruiting in a way we will see in Chapter 11 is called **convenience sampling**. Participants in research are those who are available through the local environment, usually a university, or through online systems. Since it is not logistically possible to use a perfectly diverse sample, we accept and should be aware of some limitations in the external validity of our conclusions. Within that constraint we most commonly maintain the internal validity of the research process by the simple expedient of random assignment of participants to conditions.

As we consider the practicalities of carrying out a research protocol, we should be aware of another class of extraneous variables that are difficult to fully control related to the fact that participants generally know that they are in a research study. In most experimental research, participants are fully aware that they have volunteered to participate in a psychological science project. This very likely has substantial effects on behavior for research on sensitive topics like stereotype bias but may also influence measures much more broadly.

Demand Characteristics

The general set of issues related to unexpected environmental effects on experimental methodology are termed **demand characteristics**. These reflect influence or bias accidentally imposed by details of the methodological procedure. A goal of experimental control is to prepare a rigorous research process that minimizes the risk of these effects distorting the experimental measures. Improperly controlled, demand characteristics can create confounds with the independent variable, leading to Type 1 errors.

The idea that participants might change their behavior simply because they are being watched is sometimes referred to as the Hawthorne Effect. This refers to an old, possibly apocryphal, study of industrial processes where every controlled change to the environment led to performance improvements. It was subsequently hypothesized that the workers in the studies simply put in more effort when they believed they were being observed for the research.

In controlled laboratory conditions, participants are generally going to be aware that they are participating in a research process where their behavior is observed and measured. As we will see in Chapter 8, Basic Ethics, standard ethical practice is to inform participants about the research they are participating in and obtain signed **informed consent** to participate. In some kinds of field research or non-experimental observational studies this may not be the case but for the current purposes, we consider experimental control when participants are aware of their participation.

Since participants will be aware of being in a study for all conditions of the independent variable, this aspect itself is maintained through constancy. However, we need to attend to details of the levels of the independent variable as they interact with the **expectations** of participants. When participants are in a condition that they expect to affect their behavior, they may exhibit changes due to those expectations instead of the actual effect of the independent variable.

The most common example of this is the well-known effect of **placebos** that occur in medical and clinical research. A placebo is a simulated treatment that lacks any active ingredient or element that should make it effective, and a placebo effect is a positive effect of such a treatment. Many folk remedies that seem to work—such as eating chicken soup for a cold or placing soap under the bed sheets to stop nighttime leg cramps—are probably nothing more than placebos. Placebo effects are not primarily driven by people's expectations that they will be effective. Many people are not surprised that placebos can have a positive effect on disorders that seem fundamentally psychological, including depression, anxiety, and insomnia. However, placebos can also have a positive effect on disorders that most people think of as fundamentally physiological in medical research. Placebo effects are interesting in their own right as they imply surprising interactions between psychological and physiological processes. However, they also pose a challenge in experimental control for researchers who want to determine whether a treatment works.

With two levels of an independent variable in an experimental design, we do not typically have to be specifically concerned about a placebo effect, but we do need to attend to the expectations of participants. In Experiment 1, participants were not told that the hypothesis for the experiment was that the deep encoding condition would lead to better memory. If they were told this, those participants may have more actively engaged with effortful study and expecting the hypothesis to be true might have influenced the results. This would create a significant validity problem for the study, potentially producing a Type 1 error.

As an aside, sometimes researchers become concerned about the opposite kind of expectation effect where they suspect the participants have become aware of the hypothesis and are deliberately producing behavior to ruin the experiment. This is both unlikely to occur and would also produce the less problematic Type 2 error if the hypothesis were correct. A concern like this more likely emerges from the fact that carrying out effective research can be very challenging and many well-conceived studies still do not work reliably.

The simplest way to avoid the basic expectations problem in a two-group independent samples design is to not inform the participants about the hypothesis or the other condition of the study that they are not participating in. This approach is referred to as a **single-blind** procedure and was the way we implemented our Experiment 1 here. This is an extremely common method for designing psychological research that strengthens the internal validity of the experiment by eliminating concerns about demand characteristics.

In some cases more extreme versions of disguising the research study are used to avoid expectation effects. There are a variety of ways in which **deception** is used in research procedures to keep participants unaware of the hypothesis. These can range from telling the participant that they are waiting for the next part of the procedure but their behavior is being watched surreptitiously, having participants believe they are interacting with another participant but it is really a research team member (confederate) or deliberately misleading participants about performance on a test to manipulate their emotional state.

Deception in research reflects a significant challenge to ethical research practices. It is one of the more common research techniques that can only be employed with oversight and awareness of the scientific regulatory body, the **Institutional Review Board**. In Chapter 8, we will touch on the ethical implications and common practice for balancing scientific rigor with fair and appropriate treatment of human participants. As a general rule, we prefer participants to know what they are engaging in when participating in research. Yet in any kind of blind design, even without overt deception, we usually cannot explain everything in advance, which we will see is an example of tension between ideal ethics and ideal experimental design.

Inadvertent bias in research procedures

When carrying out a research protocol using a design that keeps information about the hypothesis away from participants, some care has to be taken by the research team in how they interact with participants. There will often be a planned procedure for explaining some aspects of the study, but not all. Interactions with participants can even be scripted to help make sure the protocol is administered in the same way across all the participants.

Areas where bias can inadvertently creep into research procedures can come from expectations of the research team. Virtually all researchers want their experiments to succeed, which can lead to subtle effects like simply being more socially engaging and interacting with participants in the condition where participants are hypothesized will perform better. Many psychology experiments require a research team member to be in the room to observe behavior and in this case it is vitally important that interactions with participants are constant across conditions.

This kind of bias can even occur through assignment to conditions if consistent procedures are not implemented. If participants are assigned to conditions by the research team when they arrive to participate, more attentive and engaged participants might get assigned to the more challenging or interesting condition. This creates a confound of **sampling bias** that reduces the internal validity of they study. This kind of bias is a greater risk in health and clinical research with targeted populations than more general experimental psychology approaches. In these kinds of studies, it is particularly important to have a well-documented and carefully followed procedure for assignment to conditions, even if it is just as simple as alternating between conditions.

Some protocols require so much interaction between the research team and participants that it is impossible to be confident that all interactions will be free of any inadvertent bias. For these studies, it is necessary to use a **double-blind** procedure where the members of the research team are also unaware of what condition the participants are assigned to. These are logistically complex to employ so they only get used when there is significant bias risk. An example from medical research are clinical studies of the effectiveness of a new drug. In those studies, the pharmacy prepares numbered doses that appear identical and no knowledge of which are the treatment or control is available to the research team during the research protocol. After data collection is complete, the study is then unblinded and information about which dose was treatment or placebo is provided. Only then can the data be analyzed for treatment efficacy.

Fully double-blind procedures are rare in psychological science due to the difficulty of implementing these consistently. More common are simple procedures for assignment to conditions, combined with scripted interactions with participants before and during administration of the research protocol. However, it is not that uncommon to need to develop a special procedure for scoring some kinds of dependent variables when there is a subjective element to rating participants' behavior.

Avoiding bias in scoring

Because psychological science is often about factors that affect behavior, some studies use an operational definition that requires quantification of observed behavior. For example, we might be interested in evaluating the effect of an anxiety-reducing manipulation on public speaking performance. The dependent variable here would be an evaluation of how well the participants performed on a public speaking task, requiring a quantified judgment of that performance by the research team. Because the judges will need to make subjective decisions about the quality of performance, there would be a high risk of bias if they had full awareness of the participants' condition when making those ratings.

Whenever a subjective evaluation is part of scoring the dependent variable, it is common to use **independent raters** who carry out the scoring process without knowledge of the level of the independent variable. This requires having some members of the research team remain blind to condition but others may have full knowledge of the procedure. Keeping the raters unaware of the experimental condition avoids any bias influencing their rating by the experimental hypothesis. The raters are often trained with detailed instructions on how the scoring process is to be carried out and in some cases multiple raters are employed and scores across them combined.

In Chapter 2, we introduced the idea common to nearly all psychology studies that an abstract construct needs to be turned into a quantitative variable through the process of creating an operational definition. In some cases this can be a scale, like for self-esteem, or a performance measure like recognition memory accuracy. However, there are many ares within psychology where it takes a human being to provide an evaluation of behavior in order to carry out the operational definition. We might want to measure an aspect of emotional expression such as laughter, or rate the quality of partner interactions in a study of relationships. For any subjective judgment like this, we assume that experimenters who are aware of the design are also invested in the outcome of the study, and therefore are at risk for experimenter bias and should not be the source of the measure. In these cases, methodologically rigorous research relies on ratings provided without knowledge of experimental condition.

Design of Experiment 1

Our Experiment 1 reflects a handful of design decisions aimed to keep extraneous variables constant across the two conditions in the study: deep and shallow encoding. All participants rated the exact same set of 30 words, although the instructions for the rating varied as the independent variable. The words themselves were selected to be between 5 and 8 letters in length and to have a *written frequency occurrence* of 30-80 times per million. The characteristics of the words were kept similar to reduce variance in memory for the words chosen for the experiment.

Unless you have some experience in memory research using word lists, you might not have anticipated that the length or frequency of the stimulus would be important for the design. Knowing what potential extraneous variables are relevant to a specific study often requires some prior knowledge of research in that domain. Once the variables are identified, the technique for controlling them is straightforward: select words in a restricted range from a database of word frequency information.

In addition to the stimuli, note that the two scales used for rating the stimuli were also constructed to have 5 levels. Although it is unlikely that the specific number of levels on the scale will affect memory, it is good practice to keep as many design elements the same as possible across conditions.

In cases where the data collection for Experiment 1 are done in the classroom, we also gain the benefit of all the participants complete the study in the same conditions in terms of surrounding and time of day. When this experiment is completed by participants outside the classroom, there may be influences of outside distractions and attention that are outside of experimental control. Note that these would be examples of extraneous variables that increase variance, but do not confound the study because we have no reason to believe that either of the conditions of the independent variable would be more affected by distraction.

The design of Experiment 1 also includes 3 minutes of irrelevant trivia questions to be completed after performing the word rating and before the surprise recognition test. The time of the trivia task is kept constant across participants, but the number of questions answered and the content of the questions is not. The number and content of the questions experienced is allowed to vary randomly across all the participants in the study, potentially contributing to variance in the memory measure but not in a way that is confounded with the study conditions.

Practical considerations

Best practices for controlling extraneous variable in carrying out psychological research can lead to fairly elaborate and precise procedures for research personnel. As a consequence of this, it is very common for research procedures to be evaluated with a short period of **pilot testing** before starting formal data collection. Sometimes this can mean simply practicing carrying out the research procedures under observation of other researchers to ensure it is working as intended by the planned operational definitions. It can also mean running a small preliminary sample of participants to evaluate the procedure and scripts. It should be very clear in the overall research plan when pilot testing is underway and when that process is complete and formal data collection for the planned study starts. Pilot testing data is not intended for inclusion in published research and may often depend on knowledgeable members of the research team (or collaborating teams). This can affect demand characteristics of those participants making their behavior or performance importantly different from the main intended recruited sample.

A common feature of pilot testing of procedures is to include a measure referred to as a **manipulation check**. This is a measure that will often look like a dependent variable but is not part of the research hypothesis. For example, in a mood manipulation study using music to create positive/negative moods, participants might be asked after listening to the music to rate their mood. If mood ratings were not consistent with the independent variable (music type), we would have concern about the operational definition being used. In some research publications, manipulation check data may be included and even analyzed statistically but note that no real hypothesis is being tested. A statistically reliable effect that the music manipulation affected self-rated mood only validates the operational definition of the IV and does not lead to any general conclusion.

Pilot and preliminary testing can also be used to examine the distributional characteristics of the dependent variable. As we will see in the next chapter, our ability to draw inferences from our data will depend on observing statistically reliable effects of the IV on the DV. Poorly controlled extraneous variables may lead to high levels of variability in performance, which will show up as high variance and may indicate a need to improve experimental control in design. Accurate estimates of variance often require large participant samples, though, so this cannot always be anticipated.

Pilot testing is often very useful to identify potential statistical problems with floor effects or ceiling effects in the DV. Ceiling and floor effects occur when the dependent variable measurement range is not properly anticipated in the experimental design. For example, a floor effect will occur when a task is too difficult for participants. If participants are given a problem-solving task with the intention of the measure being the number of problems solved but nobody is able to solve any of the problems, everybody will score zero regardless of the IV manipulation (no reliable difference can be detected). Similarly, if all participants get all the answers correct, performance is at ceiling for all groups and again there is no possibility of observing a statistically reliable effect. Pilot testing is often used to verify that scores on the dependent variable will be within a range that allows for detectable influence from the independent variable so that we have some chance that our statistics will be effective.

Key Takeaways

- **Random assignment** to conditions in between-participants experiments is a fundamental element of experimental research. The purpose of this technique is to control extraneous variables so that they do not become confounding variables.
- Restricted participant sampling may reduce variance but should be used infrequently and carefully due to effects on generalizeability of findings.
- **Demand characteristics** have surprisingly large effects on behavior. Simply knowing that they are in a research study may change behavior of participants.
- **Blind** or **single-blind** designs keep the participants from knowing the full experimental hypothesis and influencing their behavior and are very commonly used.
- **Double-blind** designs mean that some members of the research team are unaware of which condition participants are in. This is used to avoid any bias in scoring the dependent variable, especially if there is any element of scoring that requires subjective judgment.
- Rigorous systematic procedures for data collection are important and contribute to research success. Written detailed scripts for the research process help manage both the influence of extraneous variables and demand characteristics on psychological research.

Exercises

Question 1

Craik & Tulving (1975) reports a series of studies examining the effect of various approaches to deep and shallow processing on memory. Review this publication and answer the following questions about specific experiments reported there comparing the procedure to our Experiment 1.

- In their Experiment 1, how many levels of the IV were used? What was the DV measure of memory?
- Their Experiment 5 is carefully designed to address what confounding alternative hypothesis? To do so, what aspect of the IV is made as constant as possible?
- In what way was their Experiment 9 similar to our in-class experiment? Identify some methodological differences

Question 2

You are doing a study at a local school. Because of the way things area scheduled, you can have one small testing room in the morning and another much larger testing room in the afternoon. If you have two treatment conditions (A and B), how can you assign subjects to the testing rooms so that the type of room will not lead to confounding your experiment?

Question 3

Dr. L is planning a large scale learning experiment. He would like to have 100 rats in one treatment group and another 100 in the other group. Because he needs so many rats, he says, “Well, I can’t test all these animals by myself. I’ll ask Dr. P. to help me. He can run the animals in the one group while I test the animals in the other group.” What is the potential problem with this approach and how would you improve the procedure to correct it?