Chapter 1: Experimental Research

Start by participating in the following psychology experiment:

|  |  |
| --- | --- |
| Qr code  Description automatically generated | Or follow the following link:  <https://tinyurl.com/Reber205> |

The following set of questions is based on the experiment but assumes some prior experience with psychological science. In our curriculum, Research Methods follows classes in Introduction to Psychology and basic Statistics. We therefore assume some familiarity with basic terms and analyses but we aim to reinforce understanding of these core ideas within the framework of what a simple experimental design looks like from the above example.

For answering the following questions it is useful as an exercise to cover the answers and try to answer the questions yourself before reading on. This will help you assess how much of the basic terminology and experimental approach you are already comfortable with. The terms will be defined in this chapter for general reference. The goal here will be to use the main terminology frequently enough that it simply becomes part of your understood vocabulary without need to look definitions up later on. The bolded terms below are ones to start becoming comfortable with.

1. What was this experiment about?

The general temptation for the answer to this question is to give a lot of detail about how your experience with the experiment relates to the underlying hypothesis. However, after just going through the experiment, you actually do not know what the experiment is about because you have not seen enough of the design. This is a typical experience for a participant in an experiment that has an **independent variable** that is manipulated **between-participants**. You only experienced one of the conditions, so the underlying hypothesis is not visible to you.

However, when we consider and evaluate research (especially published papers), we will start with this question and the answer we are looking for in this very basic question is the highest-level **construct** that gives the overall domain of the experiment. Here, that is simply “memory.”

As we will see, designing an experiment in psychology generally starts with something we are trying to learn about. In psychology, that will be a concept like memory, perception, anxiety, relationships, language, identity, etc. One of the specific challenges of experimental methods in psychology, as opposed to other areas of science (chemistry, physics, biology), is that while we intuitively understand each of those concepts, there is a significant amount of effort needed to turn that idea into things that can be used in research. That process is called identifying the **operational definition** of the construct, which is essentially, how are we going to capture that idea in a controlled study.

Answering the next questions will require being familiar with some technical terms that you may have encountered in prerequisite classes. If you are unfamiliar with the terms, they are defined below for your reference.

1. What was the **independent variable**?

To answer this question, you need some additional information. There were two different conditions used in this experiment. Half of the time, participants are given instructions to rate how much they like each word, on a 1-5 scale from “very much” to “not at all.” The other half of the participants get instructions to count how many vowels there are in each word and also make a response on a 1-5 scale.

The **independent variable (IV)** is the conditions created by the experimenter and applied to the participants. Here it is the instructions given for how to read and engage with the list of words. A more interesting question is what **construct** is this **independent variable** an **operational definition** of? What is the construct that the experimenter is manipulating in this study? The answer is “depth of encoding” which refers to how much engagement the participants have with the meaning of the words in the study list. Understanding why this is an interesting factor to manipulate will require some background reading to become familiar with the theory (which we will get to later).

Here, “depth” is an **experimental operation definition**, which refers to turning this concept/construct into **conditions** that can be applied to a research experimental design. Rating liking creates a higher level of depth by encouraging semantic engagement with the words. Counting vowels creates comparatively lower depth by focusing the participant on surface features of the word instead of meaning. The experiment is about how these conditions affect memory, which raises the next question.

1. What was the **dependent variable**?

The **dependent variable (DV)** in this experiment is a **measured operational definition** of memory, as in, how much memory did participants have of the word list after engaging with the work list in either of the experimental conditions. A measured operational definition turns a concept/construct into a quantitative number used to measure outcome. Here, the answer will be a numeric measure of performance on the recognition test that came at the end of the experimental protocol.

After going through the initial interaction with 30 words (“study phase”), you completed a short delay/distraction task based on answering trivia questions. Then you completed a recognition memory task in which you were presented with 60 works, the 30 you saw initially and 30 words that you did not see at the beginning. Note that you might be tempted to answer the question of “what is the DV?” with “the number of studied words you responded ‘old’ to on the test.” Here that is not quite correct as answering ‘old’ to all 60 words would not reflect good memory (because you called all the new words old). More accurate is to describe the DV as score on the recognition test, which we can count as the number of test items responded to correctly (old called old, new called new).

As an aside, if you are familiar with more complex memory or perceptual cognitive experiments, you might have encountered the idea of “signal detection theory” which provides some alternate approaches for quantifying these kinds of measured operational definitions that add precision and some additional elements of experimental control. Those more sophisticated approaches are not necessary for this simple experimental protocol.

1. State a hypothesis relating the **independent variable** to the **dependent variable**.

Correct answers to this question do not have to be correct about the data. Any statement relating the levels of the IV to scores on the DV are correct answers to a prompt like “state a hypothesis.” The hypothesis that actually drove the research is: *rating liking of words will lead to higher scores on the recognition test than counting vowels*. Stating the opposite, that counting vowels will increase recognition scores compared with rating liking is also a valid hypothesis, although we will see that it is not supported by the data.

Note that we could also state the hypothesis in terms of the constructs, i.e., *deeper encoding of words leads to better memory*. Technically this is not an accurate answer to the question #4 above because it references the IV and DV directly and therefore so should the answer. It is, however, an accurate theoretical statement about the research. The important point to note here is to be aware of the different levels of statements here. The data will tell us about the relationships of the variables we used in the experiment. We will then hope to draw a theoretical inference about the constructs as the scientific conclusions about the study. One of the important and unique aspects of psychological science is attending to the difference between the actual experimental data, which are based on **operational definitions**, and the theoretical conclusions, which are based on **constructs**. In this design, the operational definitions led us to use lists of words as the things to be remembered and one specific approach to what we mean by “depth of encoding.” These might be important limitations to consider about our conclusions, i.e., do they apply to non-word stimuli and/or how does depth influence other kinds of ways to measure memory?

1. What statistical test would we use to establish a reliable relationship between our independent and dependent variables that would allow us to test our hypothesis?

Since this is a simple two group design with participants randomly assigned to one condition or the other, the most appropriate statistical test would be a **two independent samples t-test**. While other more powerful approaches could certainly be used, it is generally more effective in presenting research results to use the simplest test that effectively communicates the main findings.

In this class, you will learn how to understand research from the perspective of being able to read and evaluate research findings, such as published papers. You will also learn how to prepare and communicate research findings based on an experiment and collected data. Statistics are the bridge from your numeric, quantitative data to statements about the conclusions and meaning of your study. As a results, we will talk about statistics in a somewhat different manner than you have previously discussed statistics in prior classes. Our focus will be on identifying the correct statistical tool, applying it appropriately and presenting the results following the field-standard format (based on the American Psychological Association; APA). This will then be used to support inferences drawn from the data that are associated with the constructs.

Experiment 1 will be used to illustrate the typical path from theory through experimental design, data collection and analysis. We start with **constructs** like *memory* and a hypothesis, *does deeper engagement with material lead to better memory?* These are then turned into an **experimental operational definition** (liking and vowel counting) and **measured operational definition** (recognition test). Data are collected and will be analyzed. The statistical test will be used to allow us to support (or not) a statement about whether the IV reliably affected the DV. From there we will draw a final conclusion about how we think the original concepts are related and whether the data support the original hypothesis (or not).

# Key definitions

**Experimental research**: The experimenter manipulates an **independent variable** and measures a **dependent variable** to observe whether the manipulation has an effect.

**Construct**: The high-level concepts we aim to do research about. Typically, these things we have an intuitive understanding of but that are too broad for immediate translation into specific experiment elements.

**Experimental operational definition**: a controlled method of implementing a specific definition of a construct into levels or categories that can be manipulated by an experimenter in order to create the independent variable(s) for an experiment protocol.

**Measured operational definition**: a quantitative measure of a construct, essentially turning an idea into something that can be characterized as a number.

**Independent variable**: The factor manipulated by the experimenter.

**Dependent variable**: The measurement collected by the experimenter.

**Experimental Hypothesis**: statement about the relationship between experimental variables that can be tested and importantly, falsified. If there are no data that would render a statement false, then it is not a falsifiable statement and is typically a description rather than a hypothesis. Note that hypotheses can be stated about the specific IV and DV used in an experiment but also stated separately about the constructs from which the IV and DV were operationally defined. With data, we have the most confidence about statements about the IV and DV but the goal of research is to draw inferences about the relationship among the constructs.

A useful approach for understanding the definition of something complex, like **experimental research**, is to define what isn’t experimental research. In **non-experimental research**, we also look for a relationship between an independent variable and a dependent variable, but the independent variable is not manipulated or controlled by the experimenter. For example, we could look for a correlation between your GPA and the score on the memory test in the demonstration experiment. Non-experimental research is a powerful tool for psychological science as well as fields such as epidemiology, economics and sociology. However, the methods of the design of research studies and tools for analysis of data for non-experimental methods are quite different. The current approach focuses on experimental methods first, followed by some discussion of contrasting these methodologies for general reference in Chapters 9-11.

[The following sections are from the original textbook and are still being integrated into the flow of our class]

# Experimental Research

In the late 1960s social psychologists John Darley and Bibb Latané proposed a counter-intuitive hypothesis. The more witnesses there are to an accident or a crime, the less likely any of them is to help the victim (Darley & Latané, 1968).

They also suggested the theory that this phenomenon occurs because each witness feels less responsible for helping—a process referred to as the “diffusion of responsibility.” Darley and Latané noted that their ideas were consistent with many real-world cases. For example, a New York woman named Catherine “Kitty” Genovese was assaulted and murdered while several witnesses evidently failed to help. But Darley and Latané also understood that such isolated cases did not provide convincing evidence for their hypothesized “bystander effect.” There was no way to know, for example, whether any of the witnesses to Kitty Genovese’s murder would have helped had there been fewer of them.

So to test their hypothesis, Darley and Latané created a simulated emergency situation in a laboratory. Each of their university student participants was isolated in a small room and told that they would be having a discussion about university life with other students via an intercom system. Early in the discussion, however, one of the students began having what seemed to be an epileptic seizure. Over the intercom came the following: “I could really-er-use some help so if somebody would-er-give me a little h-help-uh-er-er-er-er-er c-could somebody-er-er-help-er-uh-uh-uh (choking sounds)…I’m gonna die-er-er-I’m…gonna die-er-help-er-er-seizure-er- [chokes, then quiet]” (Darley & Latané, 1968, p. 379).

In actuality, there were no other students. These comments had been prerecorded and were played back to create the appearance of a real emergency. The key to the study was that some participants were told that the discussion involved only one other student (the victim), others were told that it involved two other students, and still others were told that it included five other students. Because this was the only difference between these three groups of participants, any difference in their tendency to help the victim would have to have been caused by it. And sure enough, the likelihood that the participant left the room to seek help for the “victim” decreased from 85% to 62% to 31% as the number of “witnesses” increased.

## The Parable of the 38 Witnesses

The story of Kitty Genovese has been told and retold in numerous psychology textbooks. The standard version is that there were 38 witnesses to the crime, that all of them watched (or listened) for an extended period of time, and that none of them did anything to help. However, recent scholarship suggests that the standard story is inaccurate in many ways (Manning, Levine, & Collins, 2007). For example, only six eyewitnesses testified at the trial, none of them was aware that they were witnessing a lethal assault, and there have been several reports of witnesses calling the police or even coming to the aid of Kitty Genovese. Although the standard story inspired a long line of research on the bystander effect and the diffusion of responsibility, it may also have directed researchers’ and students’ attention away from other equally interesting and important issues in the psychology of helping—including the conditions in which people do in fact respond collectively to emergency situations.

 The research that Darley and Latané conducted was a particular kind of study called an experiment. Experiments are used to determine not only whether there is a meaningful relationship between two variables but also whether the relationship is a causal one that is supported by statistical analysis. For this reason, experiments are one of the most common and useful tools in the psychological researcher’s toolbox. In this chapter, we look at experiments in detail. We will first consider what sets experiments apart from other kinds of studies and why they support causal conclusions while other kinds of studies do not. We then look at two basic ways of designing an experiment—between-subjects designs and within-subjects designs—and discuss their pros and cons. Finally, we consider several important practical issues that arise when conducting experiments.

# **Experiment Basics**

Learning Objectives

1. Explain what an experiment is and recognize examples of studies that are experiments and studies that are not experiments.
2. Distinguish between the manipulation of the independent variable and control of extraneous variables and explain the importance of each.

## What Is an Experiment?

As we saw earlier in the book, an experiment is a type of study designed specifically to answer the question of whether there is a causal relationship between two variables. In other words, whether changes in one variable (referred to as an independent variable) causes change in another variable (referred to as a dependent variable). Experiments have two fundamental features. The first is that the researchers manipulate, or systematically vary, the level of the independent variable. The different levels of the independent variable are called conditions. For example, in Darley and Latané’s experiment, the independent variable was the number of witnesses that participants believed to be present. The researchers manipulated this independent variable by telling participants that there were either one, two, or five other students involved in the discussion, thereby creating three conditions. For a new researcher, it is easy to confuse these terms by believing there are three independent variables in this situation: one, two, or five students involved in the discussion, but there is actually only one independent variable (number of witnesses) with three different levels or conditions (one, two or five students). The second fundamental feature of an experiment is that the researcher exerts control over, or minimizes the variability in, variables other than the independent and dependent variable. These other variables are called extraneous variables. Darley and Latané tested all their participants in the same room, exposed them to the same emergency situation, and so on. They also randomly assigned their participants to conditions so that the three groups would be similar to each other to begin with. Notice that although the words manipulation and control have similar meanings in everyday language, researchers make a clear distinction between them. They manipulate the independent variable by systematically changing its levels and control other variables by holding them constant.

## Manipulation of the Independent Variable

Again, to manipulate an independent variable means to change its level systematically so that different groups of participants are exposed to different levels of that variable, or the same group of participants is exposed to different levels at different times. For example, to see whether expressive writing affects people’s health, a researcher might instruct some participants to write about traumatic experiences and others to write about neutral experiences. The different levels of the independent variable are referred to as conditions, and researchers often give the conditions short descriptive names to make it easy to talk and write about them. In this case, the conditions might be called the “traumatic condition” and the “neutral condition.”

Notice that the manipulation of an independent variable must involve the active intervention of the researcher. Comparing groups of people who differ on the independent variable before the study begins is not the same as manipulating that variable. For example, a researcher who compares the health of people who already keep a journal with the health of people who do not keep a journal has not manipulated this variable and therefore has not conducted an experiment. This distinction is important because groups that already differ in one way at the beginning of a study are likely to differ in other ways too. For example, people who choose to keep journals might also be more conscientious, more introverted, or less stressed than people who do not. Therefore, any observed difference between the two groups in terms of their health might have been caused by whether or not they keep a journal, or it might have been caused by any of the other differences between people who do and do not keep journals. Thus the active manipulation of the independent variable is crucial for eliminating potential alternative explanations for the results.

Of course, there are many situations in which the independent variable cannot be manipulated for practical or ethical reasons and therefore an experiment is not possible. For example, whether or not people have a significant early illness experience cannot be manipulated, making it impossible to conduct an experiment on the effect of early illness experiences on the development of hypochondriasis. This caveat does not mean it is impossible to study the relationship between early illness experiences and hypochondriasis—only that it must be done using nonexperimental approaches. We will discuss this type of methodology in detail later in the book.

## Causation versus correlation

One of the principal goals of scientific research is to draw causal inferences about the domain of study. In the Kitty Genovese example, the question is whether the presence of bystanders causes people to be less likely to help. In the demonstration Experiment 1, the question is whether deeper semantic encoding when reading a list of words will cause memory for those words to be better. The power of experimental research is that by manipulating the independent variable, you create a context in which those causal inferences are valid. If we design our experiments well, avoiding confounds or other bias (how to do this is the topic of Chapters 3 and 4) then when we obtain statistically reliable results, we can confidently conclude there is a causal relationship among our variables.

In later chapters when we review nonexperimental methodologies, we will see that this is the essential challenge with those approaches: they are much more likely to produce correlations among variables of interest and we will have to work much harder to try to establish causation. Balancing that is the fact that nonexperimental approaches are often a lot more effectively grounded in the real world and more clearly applicable. Using the terminology defined above, we can say that the operational definitions of the experimental variables are much closer to the real constructs of interest. As we discuss the techniques for designing robust experimental approaches, we should always be aware that the process of turning a construct, e.g., memory, into a laboratory controlled operational definition like score on a recognition test carries the risk of giving us robust causal inference about the experimental operational definitions that might not apply to the big picture theory as well as we would like. We will return to this tension between different types of design periodically as we review methodological techniques throughout the textbook.

## Errors in Inference from Research

One important reason to pay attention to correlation/causation issues is that it is a major source of one of more common and potentially damaging errors that gets made in drawing conclusions from scientific work. It is common to categorize the two forms of possible error from science as Type 1 error and Type 2 error. A Type 1 error is a false claim where an inference is drawn from a research finding that is not true. This kind of error can cause substantial damage if policies or guidelines are built on a false finding, potentially causing the harm to spread widely. For this reason, we are generally much more concerned with Type 1 errors in research. Much of the techniques that will be discussed in the textbook are aimed at helping you to understand where these errors can occur. While we will be focusing on techniques and methodologies for carrying out research, the same critical thinking skills apply to understanding research you will encounter as published research or media reports summarizing research reports.

A Type 2 error is one where our hypothesis was true but we did not obtain data reliably supporting the hypothesis. This kind of error tends to be less damaging as many research projects that do not obtain reliable results do not get reported broadly in the scientific literature, minimizing their impact. It is still disappointing when your experiment does not support your hypothesis – although it is important to also note that sometimes your hypothesis is simply wrong. In most cases, the main cost of a Type 2 error is wasted time and effort of the experimenter and research team (there are exceptions to this, such as “glyphosate does not cause cancer” is a bad place for a Type 2 error, but this is less common in psychological research). Rigor and care in experimental design and analysis goes a long way towards minimizing Type 2 errors.

An example of the danger of a Type 1 error is the statement “vaccines cause autism.” This incorrect assertion his responsible for an immense amount of misery and death among people who mistakenly accepted it as fact. Because of the scale of damage that can occur from this kind of mistake, we need to be extremely careful as scientists about the conclusions we draw from our work. As readers or consumers of science, it is important to read scientific findings with an appropriate understanding of the methodologies, especially when we are exposed to these findings through media summaries. In this class, you will learn how to evaluate the methodology behind published claims to identify where there is a risk of a Type 1 error. Mistaking correlation for causation is one major cause of Type 1 error, particularly in non-experimental research. A related error that is more common in experimental is a confounded variable that provides an alternate explanation of the results. Others include problems with the operational definition or biased sampling of participants. Each of these issues will be discussed in depth in upcoming chapters with specific guidelines for how to plan research to minimize the risk of this kind of mistake.

## References

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility, 377–383.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., . . . Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans (12), 2512-2518. <https://psycnet.apa.org/doi/10.1093/brain/123.12.2512>

Moseley, J. B., O’Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., … Wray, N. P. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee, 81–88.

Manning, R., Levine, M., & Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of the 38 witnesses, 555–562.

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: Recent advances and current thought. 565–590.

Shapiro, A. K., & Shapiro, E. (1999). Baltimore, MD: Johns Hopkins University Press.