8 Research Ethics

It is somewhat traditional in presenting an introduction to Research Ethics to start with a litany of famous historical ethical lapses to motivate the explanation for the current standard regulatory framework within which science with human participants works. Here we will take a different approach reflecting the hands-on style of learning psychological science. First we will consider the perspective of the research participant given what we have previously discussed about rigorous experimental control. Then we will review how scientists work with and within the regulatory framework that provides oversight and ensures ethical practice.

If you have had the opportunity to participate in a psychological research study, for class credit or pay, you are likely familiar with the sense of trust you need to have towards the experimenter. As we saw in earlier chapters about experimental control, the participant is placing themselves into a situation that has been carefully and meticulously controlled by the experimenter to minimize the impact of extraneous variables. One or a very few small elements of the experience are being explicitly manipulated in order to assess effects on a dependent variable that is being quantitatively assessed from performance. Because of concerns about demand characteristics and expectation effects, the participant may not be fully aware of every aspect of behavior that is being manipulated, controlled or measured. The participant

has to accept being placed in this context and maintain trust that the experimenter will not allow them to be treated unfairly or harmed.

A major element of establishing that trust is that participants should initially be presented with an **Informed Consent** form that will tell them about the research study in which they are asked to participate. That form needs to be read, understood and signed by the participant before the procedure can begin or any data collected. A crucial element of this process is that the participants knows that their **participation is entirely voluntary** and no negative consequences of failing to participate can be in place to coerce them into research participation. Because in some cases, some aspects of the experimental hypothesis cannot be made completely clear in advance, this document may not explain every detail fully. Throughout this document will be indications, often marks or stamps, that the exact content of this form has been thoroughly examined by the **Institutional Review Board** and found to appropriately meet the balance of demands between scientific rigor and fair and ethical treatment of the participants.

The involvement of the Institutional Review Board, frequently referred to by its acronym **IRB**, should provide assurance to the participant that the study is operating within the regulatory framework that guarantees compliance with standard ethical practice. This chapter will focus on the elements of ethics related to the **fair and equitable treatment of human participants in research**. Additional topics related to the ethical practice of reporting and disseminating science will be covered when we return to this topic in Chapter 18, Responsible Conduct of Research.

As an experimenter, working to carry out a new research study, you are expected to work within this framework and with institutional oversight over your research process. The experimenter is responsible for creating the informed consent form and sending it together with a detailed summary of the research protocol to the IRB for review and approval. The submitted research protocol has to fully document all experimental procedures and the stimuli to be used in the study. The main review goal of the IRB is to carefully evaluate all the **risks to the participant** that might be incurred

through the procedure. These are then weighed against the **benefits** of carrying out the scientific study. If the benefits outweigh the risks, then the study is judged to be ethically compliant and can be carried out.

As we will see, a large proportion of studies in experimental psychology are easily seen to be **minimal risk**, which is a technical designation reflecting no likely harm could occur to the participants. Nothing bad can happen as a consequence of participating in a study like Experiment 1. However, it is important to note that the decision that a research study is minimal risk is not up to the experimenter leading the study or their research team. The research protocol still needs to be reviewed by the IRB, who provide the formal judgment of that designation.

CITI Training

If you have the opportunity to work within a research lab in the future, you will very likely be asked to complete *CITI training* and obtain certification. The Collaborative Institutional Training Initiative (CITI) is the main source of ethics training for researchers in psychology, health and medicine. Completing the training means signing up with their service, completing training and test modules and connecting that account to your university system. The training modules you will most likely encounter for psychological research are Good Clinical Practice (GCP) and Social, Behavioral, Educational (SBE) research topics. It may also include specialized content in Biomedical Research if biomedical methods like imaging or working with patient populations is part of the research protocol. In addition, all CITI training requires knowledge of Responsible Conduct of Research (Chapter 18). The university IRB then verifies your certification and allowability to be part of a research team associated with the approved research protocol.

Learning Objectives

- Describe a simple framework for thinking about ethical issues in psychological research.
- Give examples of common ethical issues that arise in psychological research that affect research participants.
- Describe the history of ethics codes for scientific research with human participants.
- Describe several strategies for identifying and minimizing risks and deception in psychological research.
- Understand informed consent and debriefing procedures, administering a consent form.
- APA ethics code for reference: https://www.apa.org/ethics/code
- The basics of ethical research are all founded on the simple idea of treating research participants with fairness and respect. They should be participating voluntarily in the research protocol with as much information about what is expected of them as can be made available. For most, basic scientific studies in psychological science, these guidelines will maintain compliance with the spirit of best practices in research ethics.

In this chapter we explore the ethics of scientific research in psychology. We begin with a general framework for thinking about the ethics of scientific research in psychology. Then we look at some specific ethical codes for biomedical and behavioral researchers —focusing on the Ethics Code of the American Psychological Association. Finally, we consider some practical tips for conducting ethical research in psychology.

Ethics is the branch of philosophy that is concerned with morality—what it means to behave morally and how people can achieve that goal. It can also refer to a set of principles and practices that provide moral guidance in a particular field. There is an ethics of business, medicine, teaching, and of

course, scientific research. As the opening example illustrates, many kinds of ethical issues can arise in scientific research, especially when it involves human participants. For this reason, it is useful to begin with a general framework for thinking through these issues.

Moral Principles

Current practice guiding ethical approaches to research are based on a short set of core principles. Five basic guidelines are listed in the APA guidance on ethics:

1. **Beneficence**
2. **Fidelity and Responsibility**
3. **Integrity**
4. **Justice**
5. **Respect for People's Rights and Dignity**

These reflect an extension of a core set of three principles from The Belmont Report, a major milestone in the development of a common standard for ethical research (see below). The Belmont Report emphasized three core principles: Beneficence, Justice, and Respect for Persons. The goal of starting with simple, intuitive principles is to emphasize that ethics arises from basic social expectations and customs. Most basic ethical questions are fairly straightforward and standard procedures effectively minimize risk of any costs while obtaining the benefits of new scientific knowledge. Later we will touch on more complex questions in research areas where substantial costs must be considered in the context of potentially important findings.

Respect for people is a core and standard idea. Your participants are participating in your research study, which is of benefit to you. Even if participants are compensated for participation, they have to trust you not to impose extreme or unfair conditions on them that are disproportionate to their expectations. As we will see below when reviewing the historical examples that led to the characterization of the current ethical regulatory framework, most of the egregious violations of ethics arose from a failure of

basic respect for human rights and dignity.

Beneficence reflects the fact that science should be done with the intention of increasing societal benefit. An implication of this idea is that our research studies should be designed to be internally valid and with the intention to publicize findings to the broader scientific community. Poorly designed research is actually unethical in that you are imposing some costs on your participants (even when these are minimal) but if no inferences can be drawn from the results because of design error, the research fails to meet the standard of beneficence.

The idea of justice in ethical research reflects the idea that science should benefit all humanity. The need to include this as a core element of our moral principles for ethical research is sadly due to the history of human experience being dominated by assumptions based on an in-group of people being more important or valuable than out-groups, an "us" versus "them" mentality. Segmenting humanity into groups in which some are more valuable is the core principle of racism (and sexism, classism, and other forms of minority oppression), which unfortunately substantially influenced some historical scientific research. Modern ideas of treating all humans equally have not always been reflected in research, which showed evidence of the cultural expectations of that time, e.g., when racism was more prevalent and/or accepted. To improve on this, we always consider the question of justice and the equity issues implied in research – does the

Be nice

A lot of the moral framework for research boils down to simply *be nice to your participants*, they are doing you the favor of being in your study. Just treat your participants as humans deserving of respect and dignity.

Of course, if everybody found it really easy to be nice to people all the time, the world would look a lot different than it does. In science, we have our regulatory oversight framework to help.

proposed work aim to benefit all humanity?

The ideas of fidelity and integrity in the APA code reflect the dangers associated with the impact of unethical scientific claims on public policy, understanding and trust. We will return to discuss these ideas in more depth in Chapter 18 in the context of Responsible Conduct of Research as a component of ethical research.

Weighing Risks Against Benefits

We start our understanding of ethical research with these principles and then we need to understand that the risks (costs) of research involve transgressing against these. As noted above, the frequent need for at least some deception is a minor violation of the concept of respect for persons. The goal of the starting principles is to guide the analysis of all of the negative aspects of a research study, i.e., the risks or costs, and then to consider these in balance with the scientific gains that can be obtained by carrying out the research.

Scientific research in psychology can be ethical only if its risks are outweighed by its benefits. Among the risks to research participants are that a treatment might fail to help or even be harmful, a procedure might result in physical or psychological harm, and their right to privacy might be violated. Risk of physical harm in psychology studies is fairly rare as many experimental procedures and manipulations do control physical aspects of behavior. As you might expect, risks of psychological distress need to be assessed and considered. Emotional or psychological distress may occur when participants are exposed to stimuli or tasks that evoke anxiety, fear, or sadness. This is particularly relevant in studies involving sensitive topics or emotional manipulations.

Among the potential benefits are receiving a helpful treatment, learning about psychology, experiencing the satisfaction of contributing to scientific knowledge, and receiving money or course credit for participating. Scientific research can have risks and benefits to the scientific community and to

society too. A risk to science is that if a research question is uninteresting or a study is poorly designed, then the time, money, and effort spent on that research could have been spent on more productive research. A risk to society is that research results could be misunderstood or misapplied with harmful consequences. Of course, the benefits of scientific research to science and society are that it advances scientific knowledge and can contribute to the welfare of society.

The regulatory framework within which the risks and benefits are weighed in order to evaluate the ethics of a psychological science study is the same framework used for all research with human participants, including health and medical research. This may be surprising given that clinical research with patients would seem to have a very different overall profile for both the risk to participants and the reward. The underlying principle of benefits outweighing risks still applies and the evaluation process for a new, proposed research study makes no assumptions about the likely risk level based on the broad domain of the research area. The primary consequence of treating all science the same initially is that researchers considering research hypotheses that are minimal risk must still document how their work does not impose any hidden or unexpected risks to participants.

Historical Overview

The rigor and care with which the scientific ethical framework is applied is motivated by historical examples where scientific work was done that clearly transgressed against these ideas. Identification of events associated with unethical research led to three major efforts in describing sets of guiding principles that would prevent recurrence of these kinds of transgressions. One of the earliest ethics codes was the **Nuremberg Code**—a set of 10 principles written in 1947 in conjunction with the trials of Nazi physicians accused of shockingly cruel research on concentration camp prisoners during World War II. It provided a standard against which to compare the behavior of the men on trial—many of whom were eventually convicted and either

imprisoned or sentenced to death. The Nuremberg Code was particularly clear about the importance of carefully weighing risks against benefits and the need for informed consent. The **Declaration of Helsinki** is a similar ethics code that was created by the World Medical Council in 1964. Among the standards that it added to the Nuremberg Code was that research with human participants should be based on a written protocol—a detailed description of the research—that is reviewed by an independent committee. The Declaration of Helsinki has been revised several times, most recently in 2004.

In the United States, concerns about the Tuskegee study and others led to the publication in 1978 of a set of federal guidelines called the Belmont Report. **The Belmont Report** explicitly recognized the principle of seeking justice, including the importance of conducting research in a way that distributes risks and benefits fairly across different groups at the societal level. It also recognized the importance of respect for persons, which acknowledges individuals' autonomy and protection for those with diminished autonomy (e.g., prisoners, children), and translates to the need for informed consent. Finally, it recognized the principle of beneficence, which underscores the importance of maximizing the benefits of research while minimizing harms to participants and society.

The Belmont Report became the basis of a set of laws—the **Federal Policy for the Protection of Human Subjects**—that apply to research conducted, supported, or regulated by the federal government. An extremely important part of these regulations is that universities, hospitals, and other institutions that receive support from the federal government must establish an institutional review board (IRB)—a committee that is responsible for reviewing research protocols for potential ethical problems. An IRB must consist of at least five people with varying backgrounds, including members of different professions, scientists and nonscientists, men and women, and at least one person not otherwise affiliated with the institution. The IRB helps to make sure that the risks of the proposed research are minimized, the benefits outweigh the risks, the research is carried out in a fair manner, and the informed consent procedure is adequate.

https://www.hhs.gov/ohrp/international/ethical-codes-and-research-standards/index.html

For psychology, additional guidance is provided by the APA's Ethical Principles of Psychologists and Code of Conduct (also known as the APA Ethics Code) was first published in 1953 and has been revised several times since then, most recently in 2016. It includes about 150 specific ethical standards that psychologists and their students are expected to follow. Much of the APA Ethics Code concerns the clinical practice of psychology—advertising one's services, setting and collecting fees, having personal relationships with clients, and so on. For our purposes, the most relevant part is Standard 8: Research and Publication and this serves as a good overall reference description of the key issues. https://www.apa.org/ethics/code

Even within these guidelines, it is not necessarily easy to weigh the risks of research against its benefits because the risks and benefits may not be directly comparable. For example, it is common for the risks of a study to be primarily to the research participants but the benefits primarily for science or society. Consider, for example, Stanley Milgram's original study on obedience to authority (Milgram, 1963). The participants were told that they were taking part in a study on the effects of punishment on learning and were instructed to give electric shocks to another participant each time that participant responded incorrectly on a learning task. With each incorrect response, the shock became stronger—eventually causing the other participant (who was in the next room) to protest, complain about his heart, scream in pain, and finally fall silent and stop responding. If the first participant hesitated or expressed concern, the researcher said that he must continue. In reality, the other participant was a confederate of the researcher—a helper who pretended to be a real participant—and the protests, complaints, and screams that the real participant heard were an audio recording that was activated when he flipped the switch to administer the "shocks." The surprising result of this study was that most of the real participants continued to administer the shocks right through the confederate's protests, complaints, and screams. Although this is considered one of the most important results in psychology—

with implications for understanding events like the Holocaust or the mistreatment of prisoners by US soldiers at Abu Ghraib—it came at the cost of producing severe psychological stress in the research participants.

This research was done at a time when the long-term implications of post-traumatic stress disorder (PTSD) were not as well understood as they are now. Although this study was not governed by systematic oversight as modern research is, the idea that the basis of the research could be explained to the participants after the study with an expectation that there would be no long-term effects might have been seen as justifying the scientific benefit. However, the modern understanding of the potential risk of unintentionally creating a long-lasting psychiatric challenge in research participants would clearly indicate that this project cannot be carried out ethically. This highlights one of the very difficult aspects of scientific ethics in that the risk/reward balance depends on current scientific understanding and this can change as science progresses.

The guidelines also reflected an understanding that researchers must conduct their research in a just manner. At a broader societal level, members of some groups have historically faced more than their fair share of the risks of scientific research, including people who are institutionalized, are disabled, or belong to racial or ethnic minorities. A particularly tragic example is the Tuskegee syphilis study conducted by the US Public Health Service from 1932 to 1972 (Reverby, 2009). The participants in this study were poor African American men in the vicinity of Tuskegee, Alabama, who were told that they were being treated for "bad blood." Although they were given some free medical care, they were not treated for their syphilis. Instead, they were observed to see how the disease developed in untreated patients. Even after the use of penicillin became the standard treatment for syphilis in the 1940s, these men continued to be denied treatment without being given an opportunity to leave the study. The study was eventually discontinued only after details were made known to the general public by journalists and activists. Subsequent investigation identified a number of additional studies in which medical treatments or interventions were not applied in a fair and

equitable manner to all participants. Most of these cases also reflected scientific mistreatment of marginalized communities. It is now widely recognized that researchers need to explicitly consider issues of justice and fairness at the societal level to avoid any possible creation of systematic mistreatment.

Who provides regulatory oversight?

All research universities and major research institutions all maintain an **Office for the Protection of Human Research Participants.** The **Institutional Review Board** is generally housed within this department. Smaller universities or hospitals engaging in research will sometimes contract with external consulting firms to provide regulatory oversight if there is insufficient local need for a whole department.

The office for research oversight consists of specially trained research staff who help manage protocols, submissions for review and any issues that arise related to compliance with ethical practice or adverse events during research procedures. The staff also support researchers who lead research protocols, but they do not participate directly as IRB panel members (large universities will also tend to have several panels with differing specializations). IRB members are mainly faculty from other departments or research domains who can evaluate proposed research protocols to verify compliance with best practices in ethical research. In addition, most IRB's have participation from community members outside the university to provide the perspective of a layperson who has not directly led research projects. These IRB members are the ones charged with applying current research rules. The department staff participate in a broader national (or international) conversation about the current understanding of best practices and how these change with new laws passed, recent court cases or the introduction of novel technologies in research.

Ethical Research in Practice

In a formal research process, all researchers involved in data collection or analysis must meet institutional training requirements for procedures and polices related to human participant research. A very common training tool is provided by CITI, which is an external online training platform that can certify training completion with the researcher's host institution. Everybody in the research process including faculty, laboratory staff, graduate and undergraduate student researchers all must meet the institutional training requirements.

Once training is certified, the researchers prepare a description of the research protocol, recruiting process and informed consent procedures to the IRB for review. The initial assessment of the IRB staff is to determine (a) does the research protocol qualify as *human participants research* and (b) what level of risk is associated with the proposal. All the experimental research described so far in this text qualifies as human participants research (protocols that do not qualify will be discussed briefly in Chapter 19) as they reflect systematic data collection from humans with an intent to disseminate the findings.

The levels of possible risk inherent in the research protocol is designated as one of three levels. **Exempt research** is the lowest level or risk and includes research on the effectiveness of normal educational activities, the use of standard psychological measures and surveys of a nonsensitive nature that are administered in a way that maintains confidentiality, and research using existing data from public sources. It is called exempt because once approved, it is exempt from regular, continuous review. A common heuristic for identifying exempt research is to ask: Would it be reasonably expected to be asked of students in a class? If so, the procedure typically reflects so little risk that the research does not require additional ongoing review from the IRB. It should be noted that only the IRB can designate a protocol as exempt. This decision cannot be made by the lead researcher even if it seems obvious the protocol should be exempt.

Expedited research poses a somewhat higher risk than exempt, but still exposes participants to risks that are no greater than minimal risk (those encountered by healthy people in daily life or during routine physical or psychological examinations). Expedited review is done by one member of the IRB or by a separate committee under the authority of the IRB that can only approve minimal risk research (many departments of psychology have such separate committees). Much psychological research is either exempt or expedited in the risk level. Research that involves deception or some significant privacy risk (e.g., asking sensitive personal questions) is typically expedited and most other protocols are designated exempt.

Finally, research that does not qualify for exempt or expedited review is greater than minimal risk research must be reviewed by the full board of IRB members. This level is common in a wide range of medical research procedures that may embed substantial risk to the participants that needs to be thoroughly evaluated and carefully monitored after approval.

The IRB review designates the risk level and review process for the research protocol and may result in suggestions or requests from the IRB panel to adjust research procedures to improve the risk/reward balance. The review process then considers the process and documentation for obtaining informed consent from research participants.

Before starting a new research project, a detailed description of the research protocol must be made available to the IRB for review in order to evaluate the risks and benefits of the research. In addition, the proposed form to collect informed consent for participants must be provided and all materials used to recruit participants into the study. When possible, all the stimuli, surveys or other research instruments should be provided or described in detail. All of this information is needed to thoroughly evaluate the risks that might be inherent in the research protocol. Once identified, these can be compared to the proposed benefits of carrying out the scientific research to test the experimental hypothesis.

Informed Consent

Informed consent means obtaining and documenting people's agreement to participate in a study, having informed them of everything that might reasonably be expected to affect their decision. This includes details of the procedure, the risks and benefits of the research, the fact that they have the right to decline to participate or to withdraw from the study, the consequences of doing so, and any legal limits to confidentiality.

A critical element of the process of obtaining informed consent is to establish that the participant is aware that **participation in research is completely voluntary**. Nobody should ever be pressured or coerced into accepting the risks of a research protocol. Even the possibility of implicit pressure is considered a potential violation of this principle, making it generally improper for employees of a laboratory to participate in even unrelated research done by that lab for fear they are doing it under the threat of consequences to their employment. Participants should also be aware that they are also free to decline to continue participation at any time during the research protocol if they feel uncomfortable for any reason. Consequences for partial participation should be specified on the consent form together with explicit reassurance that there is no penalty for withdrawing at any time during the protocol.

Although the process of obtaining informed consent often involves having participants read and sign a consent form, it is important to understand that this is not all it is. Although having participants read and sign a consent form might be enough when they are competent adults with the necessary ability and motivation, many participants do not actually read consent forms or read them but do not understand them. For example, participants often mistake consent forms for legal documents and mistakenly believe that by signing them they give up their right to sue the researcher. Even with competent adults, therefore, it is good practice to tell participants about the risks and benefits, demonstrate the procedure, ask them if they have questions, and remind them of their right to withdraw at any time—in addition to having

them read and sign a consent form.

Consent forms have a specific list of required sections that address typical questions about what is asked of participants, what compensation they will receive for participation and any risks they might bear from being in the study. Most IRB's will provide a consent form template that should be followed in preparing a consent form for a research study that will include detailed instructions on form and content. It is generally necessary to follow a provided template to ensure that the proposal's consent form meets standards and criteria for the institution's guidelines.

Recruiting participants

Researchers must also provide information on methods of recruiting participants into the study. These are typically flyers, ads or emails sent

Recruiting Materials

Information provided through ads or other recruiting methods are often the first way that participants learn about the opportunity to participate in a research study. The content and method of recruiting influences whether the sample of participants will fairly represent the population. In addition, the advertising will be aimed at generating enthusiams for participating, but must also fairly reflect the possible risks to participation.

to potentially interested participants. The review of recruiting materials is primarily aimed to ensure that participants are all volunteers who chose to engage in the research and there is no question of coercion to participate. Examples of unethical coercion include threats of loss of employment, access to medical care or other benefit to participants who choose not to participate. Nobody should ever feel forced or threatened into research participation and detailed procedures are used to ensure that any hint of coercion in recruiting are avoided.

Vulnerable populations

Special guidelines govern research on designated **vulnerable populations** which generally reflect participants who may not easily be able to establish voluntary participation in research. For example, children cannot legally attest to their participation and require a parent or guardian to sign the consent form for them. The same issue occurs in neuropsychological studies with cognitively impaired older adults, requiring a guardian signature or co-signature on the consent form. Research with vulnerable populations very commonly full board review to assess all procedural details and ensure absolute compliance with best practices.

The same idea also applies to research on incarcerated prisoners but for different reasons. Because prisoners have lost their freedom to a controlled environment, it is not expected that they can confidently assert a decision to participate or decline to participate in research. For this reason, a special set of different rules governs research with prisoners with the key requirement being that the research aims to benefit this population in a direct way (not by general benefit to society). Concerns about implicit pressure making it difficult to decline participation in research are also the reason why personnel within the laboratory conducting the research are explicitly proscribed from participating, even in minimal risk studies.

Waiver of consent

There are research situations in which informed consent is either not necessary or cannot be easily obtained within the structure of the research protocol. These include situations in which the research is not expected to cause any harm and the procedure is straightforward or the study is conducted in the context of people's ordinary activities. For example, if you wanted to sit outside a public building and observe whether people hold the door open for people behind them, you would not need to obtain their informed consent. Similarly, if a college instructor wanted to compare two legitimate teaching methods across two sections of his research methods course, he would not need to obtain informed consent from his students.

Research that might qualify for a waiver of consent must explicitly request approval for this alternate process. The request will need to be supported by an explanation of the rationale for not collecting informed consent and processes to manage any negative outcomes during the research process.

Privacy

A very common risk to be aware of in psychological research, is the implied risk to participant's privacy. Research that directly assesses personal information like sexual or other private behavior clearly carries a risk to privacy but the extent of this concern is actually substantially broader. The general guideline on privacy is that all data from research participants is kept private to only research staff and publication reports only average data that obscures any individual performance. Further, the very fact that participants were engaged in a research project is also kept confidential. In medical research, the reason for this is clear as the study may be on treatment of a disease or syndrome that the participant might not wish to have publicly disclosed. In psychological research, an example of participation privacy risk are studies on topics such as implicit racial bias which might document unexpected levels of bias in undergraduate populations. As a participant,

you might prefer not to be known to have been one of the participants in the study. Although these risks are fairly rare, the standard best practice approach is to maintain full privacy and confidentiality for all participants at all times. This poses some technical challenges for data sharing across labs that we will discuss briefly in Chapter 19.

Deception

Deception of participants in psychological research can take a variety of forms: misinforming participants about the purpose of a study, using confederates, using phony equipment like Milgram's shock generator, and presenting participants with false feedback about their performance (e.g., telling them they did poorly on a test when they actually did well). Deception also includes not informing participants of the full design or true purpose of the research even if they are not actively misinformed (Sieber, Iannuzzo, & Rodriguez, 1995). For example, a study on incidental learning—learning without conscious effort—might involve having participants read through a list of words in preparation for a "memory test" later. Although participants are likely to assume that the memory test will require them to recall the words, it might instead require them to recall the contents of the room or the appearance of the research assistant.

Some researchers have argued that deception of research participants is rarely if ever ethically justified. Among their arguments are that it prevents participants from giving truly informed consent, fails to respect their dignity as human beings, has the potential to upset them, makes them distrustful and therefore less honest in their responding, and damages the reputation of researchers in the field (Baumrind, 1985).

Note, however, that the APA Ethics Code takes a more moderate approach—allowing deception when the benefits of the study outweigh the risks, participants cannot reasonably be expected to be harmed, the research question cannot be answered without the use of deception, and participants are informed about the deception as soon as possible. This approach

acknowledges that not all forms of deception are equally bad. Compare, for example, Milgram's study in which he deceived his participants in several significant ways that resulted in their experiencing severe psychological stress with an incidental learning study in which a "memory test" turns out to be slightly different from what participants were expecting. It also acknowledges that some scientifically and socially important research questions can be difficult or impossible to answer without deceiving participants. Knowing that a study concerns the extent to which they obey authority, act aggressively toward a peer, or help a stranger is likely to change the way people behave so that the results no longer generalize to the real world.

Debriefing

Debriefing is the process of informing research participants as soon as possible of the purpose of the study, revealing any deception, and correcting any other misconceptions they might have as a result of participating. Debriefing also involves minimizing harm that might have occurred. For example, an experiment on the effects of being in a sad mood on memory might involve inducing a sad mood in participants by having them think sad thoughts, watch a sad video, and/or listen to sad music. Debriefing would be the time to return participants' moods to normal by having them think happy thoughts, watch a happy video, or listen to happy music.

The debriefing process plays an important role in studies that use deception as part of the experimental methodology. In cases with significant deception, researchers may use a secondary consent form to be completed after the research protocol. These obviously do not protect participants about their participation beforehand, but allow them to express their concern by refusing to allow their data to be included in research. Times that a participant refuses to sign a consent form at debriefing would be considered an adverse event that is required to be tracked and reported to the IRB.

Nonhuman Animal Subjects

While our course is about experimental psychology with human participants, a small amount of research within psychological departments involves the use of nonhuman animal subjects. In some cases, this work is carried out under the domain of the neighboring discipline of neuroscience, which is a scientific domain often operating within or overlapping with psychological science. There are specific and highly detailed procedures for laboratory work with nonhuman animal subjects that apply to that area of research. The core ethical approach is based on the same underlying idea of balancing the risks with the benefits of the science obtained. Obviously, these procedures are subject to the same ongoing evolution as ethical standards for human research and always aim to maintain the same awareness of current scientific understandings (e.g., of the experience of pain in animals). These guidelines are organized and administered by specialized committees trained in animal welfare and operating as an **Institutional Animal Care and Use Committee (IACUC)**. All researchers working with animals complete training on these ethical issues to ensure that these scientific studies are carried out with attention to appropriate practice.

Common Risks in Psychology Experiments

As noted earlier, the primary risks common to psychology studies are methods that have the potential to create psychological distress. That can occur by exposure to challenging materials, such as stimuli or tasks that evoke anxiety, fear, or sadness. This is particularly relevant in studies involving sensitive topics or emotional manipulation. There can even be simple frustration with difficult tasks that should be monitored as a risk.

In addition, the requirements to respect privacy and maintain confidentiality are more substantial than they may initially appear. Even in a study as low risk as our example Experiment 1, participants were randomly assigned to conditions in which one group was expected to perform significantly worse

on a memory test. If you were in that group and performed worse than your classmates, you might prefer that your personal performance not be public for others to see. Just participating in studies involving sensitive topics may risk stigmatizing participants if the nature of the research becomes known. This is particularly relevant in studies on stereotypes, mental health, addiction, or other stigmatized conditions. Studies conducted in diverse cultural contexts may unintentionally offend or harm participants if researchers are not culturally sensitive. It's crucial to consider cultural norms and values to avoid misunderstandings.

The use of deception has to be managed carefully and is usually monitored very carefully by the IRB. Mild deception by omission related to participants being blind to the underlying hypothesis are considered minimal risk. Needing to overtly mislead participants in order to observe the intended behavior will require a very strong and important scientific question to be answered to justify this imposition.

Know and Accept Your Ethical Responsibilities

As the American Psychological Association (APA) Ethics Code notes in its introduction, "Lack of awareness or misunderstanding of an ethical standard is not itself a defense to a charge of unethical conduct." This is why the very first thing that you must do as a new researcher is to know and accept your ethical responsibilities. Ultimately, you as the researcher must take responsibility for the ethics of the research you conduct.

As you design your study, you must identify and minimize risks to participants. Start by listing all the risks, including risks of physical and psychological harm and violations of confidentiality. Remember that it is easy for researchers to see risks as less serious than participants do or even to overlook them completely. For example, one student researcher wanted to test people's sensitivity to violent images by showing them gruesome photographs of crime and accident scenes. Because she was an emergency medical technician, however, she greatly underestimated how disturbing these

images were to most people. Remember too that some risks might apply only to some participants. For example, while most people would have no problem completing a survey about their fear of various crimes, those who have been a victim of one of those crimes might become upset. It is often important to seek input from a variety of people, including your research collaborators, more experienced researchers, and even from nonresearchers who might be better able to take the perspective of a participant. Once you have identified the risks, you can often reduce or eliminate many of them.

Research protocols should always be designed to take active steps to maintain confidentiality of participants. You should keep signed consent forms separately from any data that you collect and in such a way that no individual's name can be linked to their data. In addition, beyond people's sex and age, you should only collect personal information that you actually need to answer your research question. Be aware also that certain data collection procedures can lead to unintentional violations of confidentiality. When participants respond to an oral survey in a shopping mall or complete a questionnaire in a classroom setting, it is possible that their responses will be overheard or seen by others. If the responses are personal, it is better to administer the survey or questionnaire individually in private or to use other techniques to prevent the unintentional sharing of personal information.

Follow Through

Your concern with ethics should not end when your study receives institutional approval. It now becomes important to stick to the protocol you submitted or to seek additional approval for anything other than a minor change. During the research, you should monitor your participants for unanticipated reactions and seek feedback from them during debriefing. One criticism of Milgram's study is that although he did not know ahead of time that his participants would have such severe negative reactions, he certainly knew after he had tested the first several participants and should have made adjustments at that point.

Severe negative reactions to research protocols are known as adverse events and should be reported to the IRB. It may similarly be necessary to track the number of participants who elect not to participate after reading the informed consent form, or who chose to withdraw after starting to participate. These events may indicate that the research protocol bears more risk than the initial review indicated, and this may in turn require adjustments to the research protocol.

Sometimes mistakes are made in carrying out the procedures associated with the research project. These are technically termed protocol violations and might reflect mistakes or misunderstandings in obtaining informed consent or accurately informing participants about research requirements. Protocol violations should always be reported to the research oversight staff associated with the IRB or office of research protection. Researchers might hesitate to report mistakes for fear of consequences up to and including having the research lab (or even department) shut down entirely, preventing all future research from continuing. However, these events should always be reported promptly and thoroughly as the consequences of minor mistakes are rarely substantial. A typical response process to an error would be to document the error, identify and correct any harm and document improvement in procedures or training to prevent future errors. In general, the IRB strongly prefers hearing about procedural errors from researchers as early as possible. When a problem is raised to them from a participant who has concerns about a negative experience in research (or worse, legal representation), this situation is almost always a much worse problem and requires much more substantial corrective action.

Evolution of ethical practice

The idea that the standards for ethical practice change over time is sometimes viewed with surprise by researchers engaging systematically with research ethics training for the first time. On reflection, this should not be surprising. Not only to societal and cultural expectations about ethics and morality

change over time, but technology changes and with it, new issues with respect to concerns such as privacy become relevant. Widespread access to communication over the internet has opened up new possibilities for research, but also raised additional questions about privacy and standard assurances related to informed consent, e.g., how confidently can we establish the identity of participants who are solely interacted with via the internet.

A minor example of this evolution over time is the preference for the word *participant* for humans who engage in research. Older publications and writing will generally refer to participants as *subjects* but in modern parlance, subjects are used exclusively for nonhuman animals. This was done acknowledging that the idea of being "subject" to experimenter control might be considered objectionable and to focus more directly on the need for humans in research to be participating voluntarily.

A more significant example has to do with an understanding of how to handle difficult questions of participant privacy. In general, we attempt to fully respect everything about participation in a research study and to maintain privacy of all information about participants. Research might be based on asking participants about personal, high-risk or even illegal behavior. This kind of research is handled specifically with minimizing the risk to the participants in mind. However, more recently the question of how to handle accidental awareness of potential issues separate from the research protocol, specifically, concern over mistreatment of minors participating in research. Most IRB's have decided that the need to report these concerns outweighs the maintenance of privacy and much developmental psychology research uses a *mandatory reporting* guideline where any such concerns are reported to oversight agencies.

Tension between Ethics and Science

Science is intended to reflect a gain in knowledge for the benefit of humanity. As described in the APA Ethics Code: "Psychologists are committed to increasing scientific and professional knowledge of behavior and people's

understanding of themselves and others and to the use of such knowledge to improve the condition of individuals, organizations, and society." In experimental research, we observe or measure people's behavior in controlled or manipulated conditions to carry out this gain of knowledge. However, subjecting people to controlled conditions and measuring their responses is fundamentally not really a nice thing to do to them. When participants are unaware of experimental conditions, there is at least a minimal aspect of deception in the research protocol, a "lie" of omission, and some protocols use considerably more deception. When we observe or measure participants' behavior, we are at some level invading their privacy. The manipulation of experimental conditions may very well put participants in a challenging situation not directly of their own choice. We consider these impositions both costs of doing research and also acknowledge the risk of negative consequences created by the research protocol.

The research process therefore cannot be carried out without some cost imposed on our research participants. It is therefore necessary to always consider the costs and benefits of each specific research study and ensure that the value of carrying out the research exceeds the costs imposed on the participants. The regulatory framework for this process is centered around the Institutional Review Board (IRB) that reviews and provides oversight of research processes. For research with human participants, there is almost always a process of obtaining informed consent from participants before they engage with the research protocol.

The regulatory framework in place to ensure that scientific work maintains appropriate ethical balance between risks and rewards applies to all kinds of research done at a university or research institution. The same review and evaluation process applies to all medical research, human research, animal research, community research, epidemiological, economic, and public health studies. The majority of psychological science research actually bears fairly little risk or cost to participants, meeting a technical consideration termed minimal risk. However, the review and evaluation processes follow a common procedure across all types of research, which is often a surprise to

beginning researchers in psychology. Practically it means that the oversight paperwork process involves a lot of questions that do not appear to directly apply to simple psychological experiments. The rationale for this approach is based on the historical occurrences of unethical research that highlight the potential problem that scientists in the past have made substantial mistakes in understanding and applying a proper understanding of the risk and costs being imposed on participants. As a consequence, the IRB, as an oversight committee, is required to evaluate the potential risks of all research studies from basic principles and not simply trust that the lead researcher on a project will carry out the research ethically.

Key Takeaways

- A wide variety of ethical issues arise in psychological research. Thinking them through requires considering how each of four moral principles (weighing risks against benefits, acting responsibly and with integrity, seeking justice, and respecting people's rights and dignity) applies to each of three groups of people (research participants, science, and society).
- Ethical conflict in psychological research is unavoidable. Researchers must think through the ethical issues raised by their research, minimize the risks, weigh the risks against the benefits, be able to explain their ethical decisions, seek feedback about these decisions from others, and ultimately take responsibility for them.
- There are several written ethics codes for research with human participants that provide specific guidance on the ethical issues that arise most frequently. These codes include the Nuremberg Code, the Declaration of Helsinki, the Belmont Report, and the Federal Policy for the Protection of Human Subjects.
- The APA Ethics Code is the most important ethics code for researchers in psychology. It includes many standards that are relevant mainly to clinical practice, but Standard 8 concerns informed consent, deception, debriefing, the use of nonhuman animal subjects, and scholarly integrity in research.
- Research conducted at universities, hospitals, and other institutions that receive support from the federal government must be reviewed by an institutional review board (IRB)—a committee at the institution that reviews research protocols to make sure they conform to ethical standards.
- Informed consent is the process of obtaining and documenting people's agreement to participate in a study, having informed them of everything that might reasonably be expected to affect their decision. Although it often involves having them read and sign a consent form, it is not equivalent to reading and signing a consent form.
- It is your responsibility as a researcher to know and accept your ethical responsibilities.
- You can take several concrete steps to minimize risks and deception in your research. These include making changes to your research design, prescreening

to identify and eliminate high-risk participants, and providing participants with as much information as possible during informed consent and debriefing.

- Your ethical responsibilities continue beyond IRB approval. You need to monitor participants' reactions, be alert for potential violations of confidentiality, and maintain scholarly integrity through the publication process.

Exercises

Questions

1. What is the main goal and purpose of the Institutional Review Board?
2. Before participating in a research experiment, all participants should generally read and sign an informed consent form. What are 3 main goals of this process intended to maintain ethical standards for the scientific work?
3. In a study of attitudes about extramarital affairs, a researcher finds that an acquaintance has participated in the study and reports having cheated on their spouse. They then find themselves torn about whether to report this information to the spouse. What kind of research ethics problem has occurred here? What research procedures should be used to keep this from occurring?

Additional Questions

- Practice: Imagine a study testing the effectiveness of a new drug for treating obsessive-compulsive disorder. Give a hypothetical example of an ethical issue from that could arise in this research.
- Discussion: It has been argued that researchers are not ethically responsible for the misinterpretation or misuse of their research by others. Do you agree? Why or why not?
- Practice: Read the Nuremberg Code, the Belmont Report, and Standard 8 of the APA Ethics Code. List five specific similarities and five specific differences among them.

- Discussion: In a study on the effects of disgust on moral judgment, participants were asked to judge the morality of disgusting acts, including people eating a dead pet and passionate kissing between a brother and sister (Haidt, Koller, & Dias, 1993). If you were on the IRB that reviewed this protocol, what concerns would you have with it? Refer to the appropriate sections of the APA Ethics Code.
- Discussion: How could you conduct a study on the extent to which people obey authority in a way that minimizes risks and deception as much as possible? (Note: Such a study would not have to look at all like Milgram's.)
- Practice: Find a study in a professional journal and create a consent form for that study. Use a standard informed consent template as a guide.

9 Factorial Design

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

Cleanliness and moral judgments

Simone Schnall and her colleagues carried out a series of simple studies examining an interesting effect where, in which they found that priming the idea of cleanliness (Experiment 1) or washing one's hands (Experiment 2) led

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

A diagram of the design helps with understanding how the two independent variables or factors here are affecting participants' behavior.

		Tested in Messy Room	
		No	Yes
Private Body Consciousness	Low	Typical moral judgments	Typical moral judgments
	High	Typical moral judgments	**Very harsh moral judgments**

There are effectively four conditions in this study. There are participants high in private body consciousness (PBC) tested in a messy room, high PBC tested in a clean room, low PBC tested in a messy room and low PBC tested in a clean room. In this chapter we will use diagrams like the one provided here to see how studies like this are organized and carried out. In the next chapter, we will use these diagrams to present data from each of the conditions and review how to interpret the results of studies with this kind of design.

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

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

Learning Objectives

1. Understanding one factor designs with more than two levels of the independent variable

2. Explain why researchers often include multiple independent variables in their studies.

3. Define factorial design and use a factorial design table to represent and interpret simple factorial designs.

4. Understand the core hypotheses embedded in a factorial design: main effects and interactions among effects

One Factor Design

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

As an example, consider a hypothetical design where participants listed to one of three kinds of auditory input while performing a spatial cognition task. The type of music is the experimental factor and the three levels are classical music, electronic dance music, and soothing ocean sounds. Participants completed as many problems as they could in 10 minutes on a "paper cutting and folding" test. Just as in prior designs, the hypothesis is that the type of sounds listened to would influence the score on the test. However, it should be clear that there are already more outcomes to consider than we would have with a 2-group design. With a 2-group design, either the independent variable affects the dependent variable (test scores) or it does not. With three groups, the null hypothesis is that the sounds have no effect and that all three conditions are essentially identical. But we can reject that

null hypothesis if any of the 3 groups shows different performance on the test from the others. The statistical tool to carry out this type of inferential statistic is the Analysis of Variance (ANOVA) which will be discussed in Chapter 10. This analysis provides a p-value that indicates the probability of the data occurring under the null hypothesis and if less than .05, we can conclude the different levels of the factor affected the dependent variable score.

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

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

The potential for problem in getting a strong conclusion from this kind of design makes factors with many levels less common in psychological research than just using two levels. Very complex designs with many levels on the factors do get used but often in specific cases with very strong theoretical

foundations and in conjunction with more complex analytical tools. All the statistical tools that will be described in this class are simplified cases derived from a more general approach based on general linear models. Extrapolating to these more complex types of analysis is beyond the scope of this text.

Factorial design

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

- The Effect of Age and Divided Attention on Spontaneous Recognition
- The Effects of Temporal Delay and Orientation on Haptic Object Recognition
- Opening Closed Minds: The Combined Effects of Intergroup Contact and Need for Closure on Prejudice
- Effects of Expectancies and Coping on Pain-Induced Intentions to Smoke
- The Effects of Reduced Food Size and Package Size on the Consumption Behavior of Restrained and Unrestrained Eaters

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

interesting effects to study in psychological research but also ones that can pose more difficulties in drawing accurate inferences from.

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

		Cell Phone	
		No	Yes
Time of Day	Daytime		
	Nighttime		

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

In principle, factorial designs can include any number of independent variables with any number of levels. For example, an experiment could include the type of psychotherapy (cognitive vs. behavioral), the length of the psychotherapy (2 weeks vs. 2 months), and the sex of the psychotherapist (female vs. male). This would be a 2 × 2 × 2 factorial design and would have eight conditions. The table below shows one way to diagram this design. In practice, it is unusual for there to be more than three independent variables with more than two or three levels each. This is for at least two reasons: For one, the number of conditions can quickly become unmanageable. For example, adding a fourth independent variable with three levels (e.g., therapist experience: low vs. medium vs. high) to the current example would make it a 2 × 2 × 2 × 3 factorial design with 24 distinct conditions. Second, the number of participants required to populate all of these conditions (while

maintaining a reasonable ability to detect a real underlying effect) can render the design unfeasible. For a 2 x 2 design, we might determine that we need 20 participants in each of the four conditions to have adequate statistical power (sensitivity to detect an effect, discussed in more detail in Chapter 12) for a total of 80. In this hypothetical 2 x 2 x 2 design diagrammed below, we would need to double that number to 160. In general, increasing complexity in factorial designs increases the number of participants required exponentially, making some complex designs essentially infeasible. As a result, we will primarily focus on designs with two independent variables. The general principles discussed here extend in a straightforward way to more complex factorial designs.

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

Assigning Participants to Conditions

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

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

Since factorial designs have more than one independent variable, it is also possible to manipulate one independent variable between participants and another within participants. This is called a mixed factorial design. For

example, a researcher might choose to treat cell phone use as a within-participants factor by testing the same participants both while using a cell phone and while not using a cell phone (while counterbalancing the order of these two conditions). But they might choose to treat time of day as a between-participants factor by testing each participant either during the day or during the night (perhaps because this only requires them to come in for testing once). Thus, each participant in this mixed design would be tested in two of the four conditions.

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

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

Non-Manipulated Independent Variables

In many factorial designs, one of the independent variables can also be a non-manipulated independent variable. In this case, the researcher measures but does not manipulate the factor and is often a characteristic that varies across participants. The study by Schnall et al. (2008) is an example of this that incorporated the participants rating of their "private body consciousness" in the design. Scores on a measure of this characteristic were used to assign participants to either high or low "condition" on this measure. In design of

this kind of factor, it is necessary to have a plan for the distinction of the rating scale into the high/low categories. This can be done by using prior research with the scale provide definitions of the categories. It can also be done by using a median split of participants. Since the median value in a group is defined as the number that splits the groups into two equal halves, this technique is guaranteed to give equal sized samples across the two levels of this factor.

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

In considering this type of design, it is important to remember that when non-manipulated independent variables are participant variables, they are by definition between-participants factors. These variables are generally assumed to be static, which is why they are measured instead of manipulated (unlike mood, for example). As long as one independent variable is manipulated, the design is still considered an experimental design overall, no matter how many other non-manipulated factors are included. However,

conclusions about the non-manipulated variables need to incorporate the fact that these were not manipulated. We would want to avoid statements such as "high private body consciousness caused harsh moral judgment in a messy room" because it implies a causal effect on a variable that was not controlled. We would prefer to state the conclusion as a "messy room caused harder moral judgments in participants with high private body consciousness." As we will review in Chapter 16, non-experimental relationships among variables are more difficult to interpret due to needed to consider and attempt to rule out alternate explanations.

Hypochondria and Memory for Health-related Words

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

Non-Experimental Studies With Factorial Designs

Thus far we have seen that factorial experiments can include manipulated independent variables or a combination of manipulated and non-manipulated independent variables. But factorial designs can also include only non-manipulated independent variables, in which case they are no longer experiments but are instead non-experimental in nature.

Consider a hypothetical study in which a researcher simply measures both the moods and the self-esteem of several participants—categorizing them as having either a positive or negative mood and as being either high or low in self-esteem—along with their willingness to have unprotected sexual intercourse. This can be conceptualized as a 2 × 2 factorial design with mood (positive vs. negative) and self-esteem (high vs. low) as non-manipulated between-participants factors. Willingness to have unprotected sex is the dependent variable. But because neither independent variable in this example was manipulated, it is a non-experimental study rather than an experiment. This is important because, as always, one must be cautious about inferring causality from non-experimental studies because of the directionality and third-variable problems. Directionality is a challenge in some non-experimental designs where you are not sure which factor happened first. For example, willingness to have unprotected sex could be affecting mood or participant's rating of self-esteem so we do not necessarily know the direction of causality. There may also be additional extraneous variables that are causing all of these measures to increase together, creating the apparent correlation.

Hypotheses in Factorial Designs

The primary goal of using a factorial design is to look for interactions among the design factors. An interaction is defined as one of the design factors modifying the effect of another design factor. For example, in the very first experiment diagrammed above, the effect of a cell phone on driving quality

might be moderate during the daytime, but much larger at night. We would then say that the time of day influences the effect of the cell phone. Factorial designs are always designed to explore the interaction of factors. If we simply wanted to look at the effect of cell phones on driving and time of day on driving independently, we would run two parallel studies that each had a simpler 2-condition design.

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

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

Key Takeaways

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

Exercises

Question 1

- After watching a group of nursery school children, we get the idea that some toys are more popular with children than others are. We would like to test the difference in time spent playing with toys that are used for building (e.g. blocks) and toys that are not (e.g., stuffed animals). Since there are many differences between boys and girls, we would also like to look at gender as an independent variable.
- Outline a factorial design for this study and describe the operational definitions of the factors (independent variables) and dependent variable. Speculate about hypotheses for what you might see if you ran this study (effect of toy, gender and any interaction).

Additional Questions

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

10 Interpretation of Factorial Data

Even though factorial designs are a relatively small conceptual increase in the complexity of an experimental protocol, they can pose surprisingly difficult challenges when trying to make sense of the data. Even experienced researchers can make mistakes in characterizing the effects observed and drawing inferences from factorial results. In this chapter, we will describe a systematic process of working through factorial data examining the main effects and then any interactions among the factors. Data visualizations are very helpful in providing an overview and with some practice, common outcome patterns can be recognized as visual patterns in data graphs.

In addition, we will introduce the statistical concept of **effect size** to help describe interaction effects. As we will see later, modern psychological science is working to incorporate improved models of statistical inference and shift away from a reliance on a simple rejection of the null hypothesis by the familiar standard $p<.05$. Here we will examine the unstandardized effect sizes, which are the simple difference in the mean performance across conditions. These will help to identify types of interaction effects: super-additive, 3:1, and crossover. The statistical tool of a post-hoc t-test is used to assess specific contrasts between conditions within a factorial design.

Learning Objectives

1. Distinguish between **main effects** and **interactions** in the results from a factorial design and recognize each.
2. Understand factorial data tables by looking at both individual conditions as **cells**, and **marginal means.**
3. Interpret and understand bar graphs and line graphs showing the results of studies with factorial designs.
4. Understand and know how to describe basic types of interactions: **super-additive**, **3:1**, **crossover**.
5. Know the role of **post hoc t-tests** to further characterize data and test targeted hypotheses.

Main Effects

In factorial designs, the main hypotheses are tested as main effects and interactions. A main effect is the effect of one independent variable on the dependent variable—averaging across the levels of the other independent variable. Common patterns of data are illustrated here with both means tables and figures to illustrate the results.

In a means table, the average performance of participants in each of the experimental conditions is shown separately, typically with means and the standard deviations shown below in parentheses. Note that in APA format, tables are accompanied by a table note indicating what the parenthetical numbers are.

In the tables and graphs below, we have an abstract design based on two factors, Factor 1 and Factor 2. Factor 1 has two conditions, A and B. Factor 2 has two conditions, X and Y. Each participant in the study is in one of the four possibilities: AY, BY, AX, BX. For illustration, we will assume a dependent variable that is scored on a 1 to 10 scale and that 20 participants were run in each of the 4 conditions in a fully between-participants design. In Chapter 9,

we reviewed a variety of examples of manipulations and participant variables that can be used to test hypotheses in 2 x 2 designs. Here, we will keep the design specifics abstract to focus on how to interpret quantitative data.

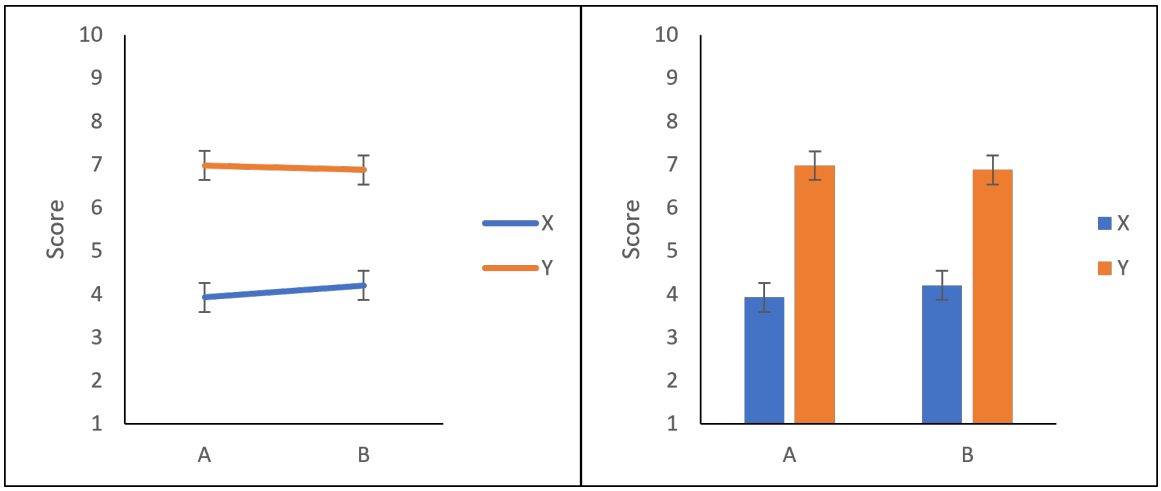
Six simulated data outcomes will be shown, each with both a **means table** and two figures illustrating the results. A means table includes the mean performance from all the participants in that specific condition of the design, typically with a measure of variance underneath in parentheses. The bottom row shows the **marginal means** across Factor 1 conditions, combining across the two levels of Factor 2, essentially setting Factor 2 aside. The very rightmost column shows the marginal means of Factor 2, combining across Factor 1 levels.

Each dataset is shown graphed as both a line graph and a bar graph. In general, experimental reports will never include both of these as they are effectively completely redundant. However, line graphs and bar plots emphasize different elements of the differences across factors slightly differently. When preparing a research report that includes a figure to illustrate the results from a factorial design, the decision of which kind of plot to use should reflect a choice of the graph that most effectively communicates the results. For these simulated data, both figures types are shown so that these differences are concretely visual in order to help with that choice.

Simulated Data 1

In the first simulated dataset, one main effect (Factor 2) is reliable, the other main effect is not affecting the DV and no interaction between the factors occurred.

Means Table		Factor 1		
		A	B	Mean
Factor 2	X	3.27 (1.51)	7.09 (1.35)	5.18 (2.39)
	Y	3.65 (1.08)	7.44 (1.44)	5.55 (2.28)
	Mean	3.46 (1.33)	7.26 (1.41)	



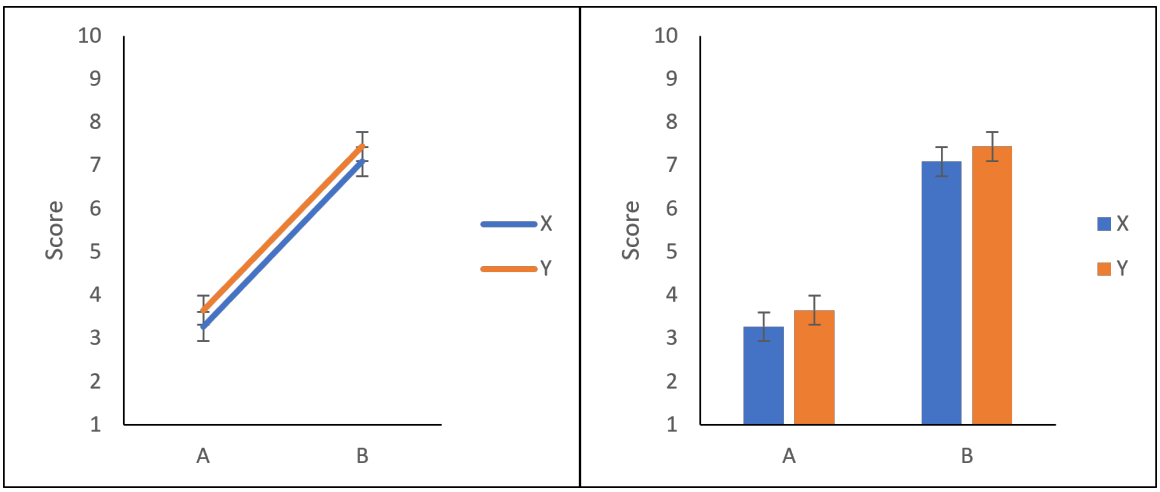
In the line graph, this pattern creates two roughly parallel, nearly horizontal lines. The distance between the lines is the effect of Factor 2. The fact that they are flat (horizontal) is reflecting the absence of an effect of Factor 1. The magnitude of the Factor 2 difference is also seen in the marginal means for Factor 2, the rightmost column of the data table (M = 4.07 for X, M= 6.93 for Y). The bar plot also clearly shows the Y condition (orange) producing higher values than the X condition (blue).

With the outcome being an effect on just one factor, we can also see the magnitude of this effect in the marginal means. The score is about 3 points higher for the Y condition than the X condition. The uncorrected effect size for Factor 2 in these data is an increase in the DV of 2.86. Later in Chapters 12 and 16 we will discuss how to use effect sizes for more sophisticated assessment of experimental effects and how to incorporate these into predicting the power and sensitivity of our designs to our hypothesized findings.

Simulated Data 2

In the second simulated dataset, one main effect (Factor 1) is reliable, the other main effect is not affecting the DV and no interaction between the factors occurred. In the line graph, this pattern creates two nearly overlapping lines that slope up the same way. The increase across the graph (from A to B) reflects the effect of Factor 1. The lack of vertical separation is due to the non-effect from Factor 2. The magnitude of the Factor 1 difference is also seen in the marginal means for Factor 1, the bottom row of the data table (M = 3.46 for A, M= 7.26 for B). The bar plot also clearly shows the B condition (right 2 bars) producing higher values than the A condition (left 2 bars).

Means Table		Factor 1		
		A	B	Mean
Factor 2	X	3.93 (1.43)	4.21 (1.22)	4.07 (1.33)
	Y	6.98 (1.28)	6.88 (1.37)	6.93 (1.33)
	Mean	5.45 (2.04)	5.55 (1.86)	



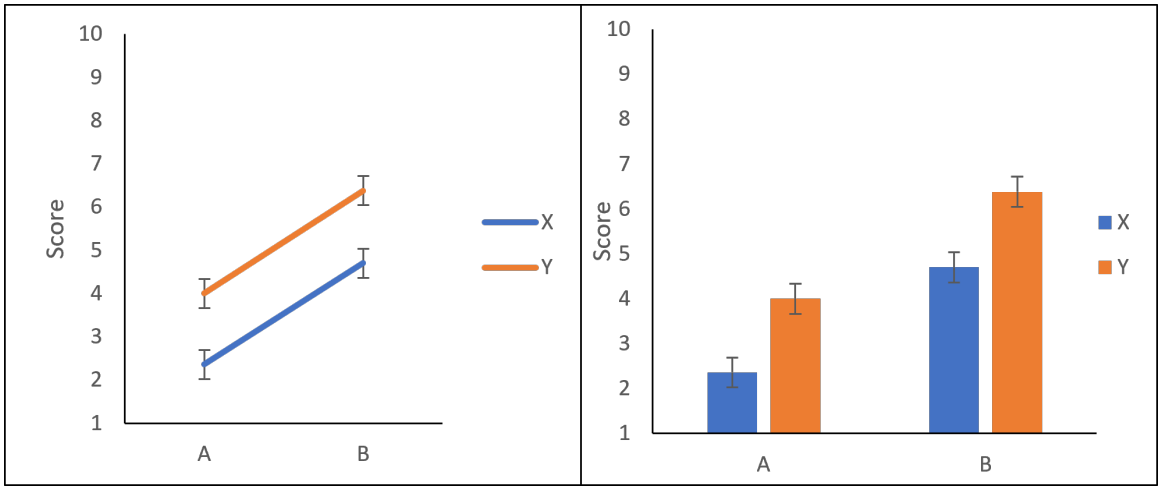
The results of this simulation are essentially the same as the first data set. We could easily transform the data by simply relabeling the factors, there are no specific rules telling you which of your independent variables in your design are Factor 1 or Factor 2. The point of providing both graphs is to indicate that when there is one reliable main effect on the dependent variable in a factorial design, you can see that in one of two ways on a line graph. It can be visually presented as two separated, roughly flat parallel lines as in the first simulation, or two nearly overlapping, steeply sloped lines as in this second graph. For bar plots, there will always be two higher bars and two lower bars reflecting the mean differences affected by that factor.

We can also still characterize the effect of Factor 1 by its uncorrected effect size seen in the marginal means. Condition B is scoring an average of 3.8 points higher than condition A.

Simulated Data 3

In this third simulated data set, both main effects are reliable but there is no interaction between them. This pattern is often mistaken for suggesting an interaction between the factors but there is none. The highest performing condition (BY) is showing the effects of both Factor 1 and Factor 2 additively.

Means Table		Factor 1		
		A	B	Mean
Factor 2	X	2.36 (1.54)	4.70 (1.54)	3.53 (1.93)
	Y	4.00 (1.08)	6.38 (1.44)	5.19 (1.84)
	Mean	3.18 (1.63)	5.54 (1.74)	



In the line graph, this pattern creates two parallel, separated lines with the same slope. As we will see below, differing slopes on a line graph is a useful visual signal of the occurrence of an interaction between factors. The increase across the graph (from A to B) reflects the effect of Factor 1. The vertical separation is due to the effect of Factor 2 (from X to Y). The marginal means

for both factors show the magnitude of the two effects independent of each other (note that these would not be independent if there was an interaction). The bar plot also shows both effects but does not imply the parallel slopes quite as easily as the line graph.

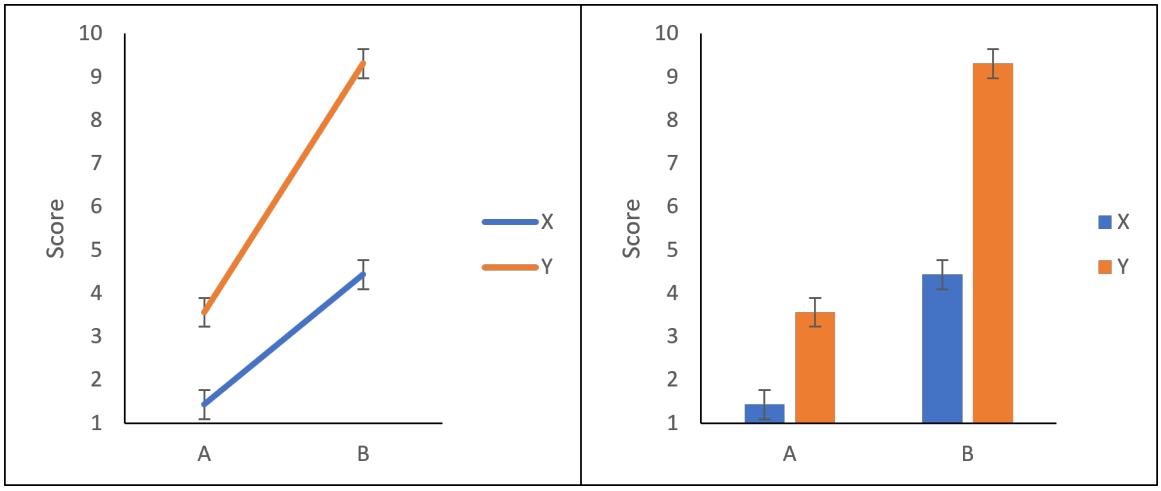
Once again looking at the marginal means for uncorrected effect sizes we see the Factor 1 effect is 2.36 points higher from A to B. The effect for Factor 2 is 1.66 points higher from X to Y. The highest scoring condition, BY, is larger than the lowest scoring condition, AX, by roughly the sum of these two effects. This is synonymous with saying there is no interaction, i.e., that the two main effects simply add together when both factors are present.

The idea that each factor can affect the DV is the basic meaning of what a main effect is and is the same way we discussed effects of our IV on the DV in the simpler 2-group designs in Chapters 2-5. With two factors, we now have two things that can affect the DV. Technically, the question of whether these factors interact with each other is mathematically the same as whether you can simply add them up to describe the data, that is, are they additive? You can see that they are additive here in the means table. The line graph also provides an important visual element to help see that. The fact that the lines are parallel is an easy way to see that the factors are combining additively. In the next three examples, we will see that the line slopes differ whenever there is an interaction. This is a visual indication that one of the factors is affecting how the other factor influences the DV, which is the definition of observing an interaction among the factors in the design.

Simulated Data 4

In these data, we have two main effects and a **super-additive interaction**. The effect of both Factor 1 and Factor 2 are to increase scores on the DV. In addition, performance in the combined condition (BY) is higher than would be predicted if the two factor effects summed together.

Means Table		Factor 1		
		A	B	Mean
Factor 2	X	1.43 (1.32)	4.43 (2.00)	2.93 (2.26)
	Y	3.56 (1.39)	9.31 (1.38)	6.43 (3.19)
	Mean	2.50 (1.72)	6.87 (2.98)	



We can describe this effect as saying the effect of Factor 1 was particularly strong in the Y condition of Factor 2. It would also be correct to describe this effect as the effect of Factor 2 was particularly strong in the B condition of Factor 1. This is another example of the symmetry of factorial designs. The design does not inherently prioritize one factor over another and there are

usually at least two equivalent ways to describe the results.

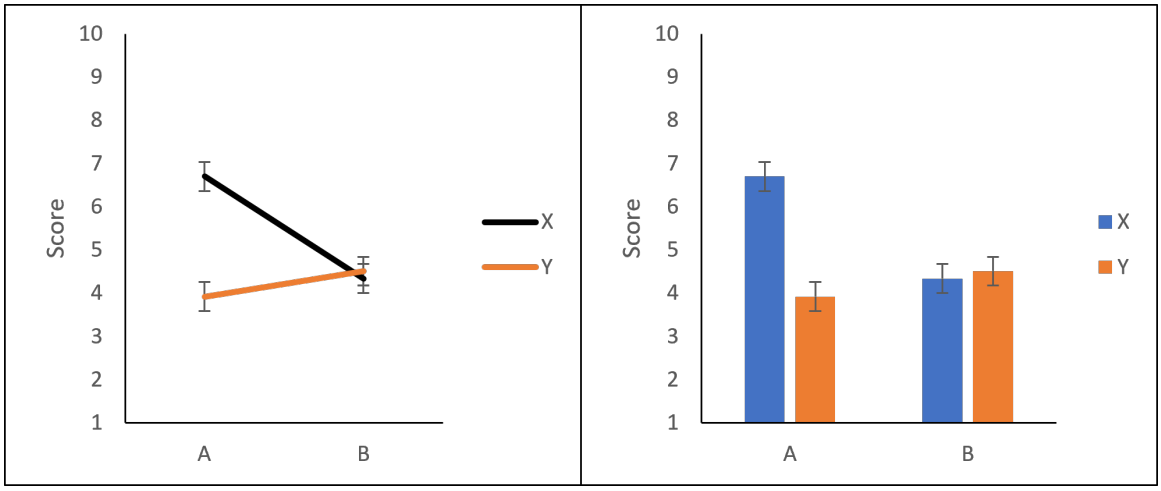
In the line graph, this pattern creates two separated lines that do not have the same slope. The different slopes are what visually signal the presence of an interaction. The main effects are visible in the same way as previous graphs. The left to right increase reflects Factor 1; the separation in lines reflects Factor 2. The marginal means for both factors show an estimate of the effects somewhat independently of each other but note how the BY condition stands out from the marginal means to show the super-additivity. The bar plot also shows both effects but the difference in slopes is again not quite as visible as in the line graph.

Compare the different here in the BY condition to the AX condition from Example 3 above. Here the BY condition is scoring even higher than the two main effects would predict independently. This kind of interaction is described as super-additive because scores in the BY condition are greater than the sum of the main effects A/B and X/Y. This interaction reflects something additional pushing up the DV score when both factors are present simultaneously that is different than either factor in isolation. The *something additional* is usually the goal in a 2x2 design and aimed to learn something new about the component variables that requires measurements across manipulations of both.

Simulated Data 5

A common data pattern in 2x2 designs is a **3:1 interaction**. In this case, one of the conditions is producing a different score from the others, which are all roughly similar (e.g., 1 high score, 3 low scores). We could describe this result using language similar to the super-additive interaction by saying the effect of Factor 2 had a large effect in the A condition of Factor 1 but little or no effect in the B condition.

Means Table		Factor 1		
		A	B	Mean
Factor 2	X	6.70 (1.76)	4.34 (1.58)	5.52 (2.04)
	Y	3.92 (1.42)	4.51 (1.29)	4.21 (1.39)
	Mean	5.31 (2.11)	4.43 (1.45)	



Technically this is just the opposite of the super-additive case (sub-additive) where the effect of Factor 2 is smaller than predicted by the main effects in the B condition. However, it is quite common to describe this result as

saying Factor 2 only matters in the A condition and does not apply to the B condition. When Factor 1 is a participants variable (e.g., an personality variable or measure like "math identification") we are observing a case where an experimental manipulation only appears to affect one subgroup of the population. This type of design and outcome are fairly common in psychological science.

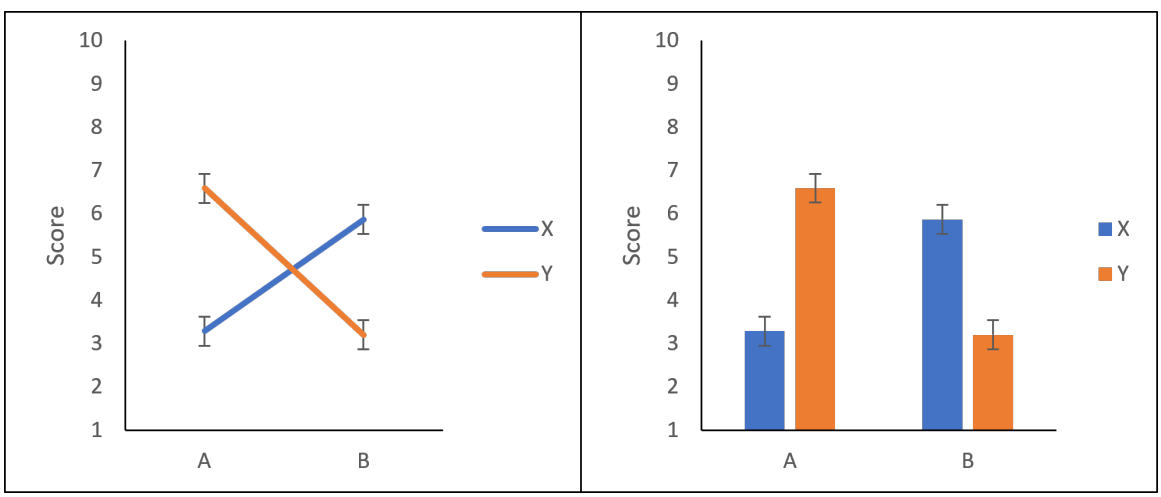
In this type of interaction we see that the bar plot graph illustrates the form of the data more effectively than the line graph. The AX condition stands out from the other three conditions which produce similar levels of performance. The one standing out from the other three is the basis for describing this as a 3:1 interaction type. We can still see that the lines are not parallel in the line graph, which reflects the fact that there is an interaction between factors.

This is also a special case in that the marginal means on the table are not very helpful in understanding the data. Both factors are showing a hint of a main effect on the marginal means, but describing the data in terms of the main effects does not necessarily contribute effectively to communicating the 3:1 interaction in the data. We could characterize this kind of effect as saying the influence of Factor 2 goes away in condition B (of Factor 1). Synonymously, we could say that the influence of Factor 2 only occurs in condition A (of Factor 1). The preferred language depends on the domain being studied and the experimental hypothesis.

Simulated Data 6

Here we see the case where the effects of the factors are essentially inverted across conditions of the other factor, which produces a **cross-over interaction**. This are somewhat complex to describe fully. We can say that the effect of Factor 2 in the A condition of Factor 1 is to increase performance while the effect of Factor 2 decreases performance in the B condition. Written descriptions tend to be wordy and these interactions are examples of where data visualizations are very valuable.

Means Table		Factor 1		
		A	B	Mean
Factor 2	X	3.29 (1.07)	5.87 (1.54)	4.58 (1.85)
	Y	6.59 (1.63)	3.20 (1.31)	4.90 (2.25)
	Mean	4.94 (2.15)	4.53 (1.96)	



In this type of interaction, we see that the line graph can makes a clear visual signal of the crossover, essentially drawing an X on the graph. It does not always create this shape as this does not occur if there is also a large main effect of Factor 2 that separates the lines. As in all interactions, the slopes of the two lines are markedly different. The structure of the data is also quite visible on the bar plot graph.

In these particular simulated data, there are no main effects of either factor as can be seen in the similar numbers across all the marginal means. This kind of data is often of theoretical importance because it suggests that neither factor has simple effects on the dependent variable and that prior work with simpler 2-group designs might have produced inconsistent or null results. As a reminder, studies that produce a null result do not establish that the IV had no effect on the DV, but only tell us that the study did not work – the IV did not affect the DV under the procedure and sampling conditions of that study. For these types of findings, we would not focus on the effect sizes characterized by the main effects. Instead, we might characterize the differences in specific conditions contrasted by use of a post-hoc t-test.

Post-hoc t-tests

The analytical tool used to identify reliable main effects and interactions is the ANOVA (analysis of variance). As we shall see, the output of the ANOVA will give us a statistical measure (the F ratio) and a p value (the probability of the data under the null hypothesis, as always) for each of the two main effects and interaction. However, the ANOVA itself does not indicate anything about the direction of the effects or the form of the interaction. For the main effects, we can use the descriptive statistics like the marginal means to support the inferential statement about differences in conditions. For the interaction term, it is often necessary to actually graph the data and visualize the form of the interaction in order to describe it accurately in the text. Nothing in the output of the ANOVA procedure indicates the form of the interaction.

In addition, for some interaction types, particularly 3:1 and cross-over interactions, it may be theoretically important to further characterize the data by targeted analysis of a subset of the data. For example, in the 3:1 interaction above, we may want to know whether the X condition is producing reliably higher numbers than the Y condition when only considering participants from the A condition of Factor 1. The tool for this is a **post-hoc t-test** on just the participants from condition A comparing scores in the X and Y conditions. This type of test is described as post-hoc because it is done after obtaining the ANOVA results and may also be described as a protected t-test. The standard analytical tool for carrying out this analysis is the T**ukey's HSD test**, which provides some additional statistical rigor for cases where multiple post-hoc comparisons may be evaluated.

For a crossover design, we can use post-hoc t-tests to evaluate both conditions separately to see if the effect of Factor 2 is reliable in just the A condition (Factor 1) and also is it reliable in the B condition. It should be noted that the fact that the interaction is statistically reliable does not automatically require that these post-hoc t-tests are individually reliable. A reliable interaction means the effects of one factor (e.g., Factor 2) are reliably

different across conditions of the other factor (e.g., Factor 1). That does not mean the effects are reliably within each condition considered separately. A reliable post-hoc t-test allows for a slightly stronger description of the results, which is often theoretically relevant.

We should also be aware when using these follow-up t-tests that there is some risk of weakening our conclusions through running too many parallel comparisons. Care needs to be taken in larger factorial designs where there can be a lot of individual conditions that we might be interested in specific comparisons between. For complex designs, we might need to consider using a Bonferroni correction for multiple comparisons or restricting our analysis to pre-specified hypotheses of specific importance.

Key Takeaways

- Main effects describe the effect of a single factor, separate from the other factors or any interaction
- Each factor in a design has the potential for a main effect. In a 2x2 design, there are two main effects.
- Main effects are visible in the marginal means in a data table.
- An interaction among factors reflects on factor modifying the effect of another factor.
- When one factor increases the impact of the other factor, it creates a super-additive interaction.
- When one factor reverses the effect of the other factor, it creates a cross-over interaction.
- When the effect of a factor is only visible in conjunction with the other factor, it creates a 3:1 interaction.
- The description and interpretation of interaction terms requires explaining the relationships among all related cells of the design, all four conditions for a 2x2

Exercises

TBA