Critical Thinking

Critical Thinking

DINESH RAMOO

THOMPSON RIVERS UNIVERSITY OPEN PRESS
KAMLOOPS, B.C.

- *An Introduction to Philosophy*, by W. Russ Payne (2015), BC Campus, is used under a CC BY-NC 4.0 license.
- *Introduction to History and Philosophy of Science*, by Hakob Barseghyan, Nicholas Overgaard, & Gregory Rupik (2018), Open Library, eCampus Ontario, is used under a CC BY 4.0 license.

Chapter 2

- Absolute knowledge. In *Introduction to History and Philosophy of Science* (Barseghyan, Overgaard & Rupik, 2018), Open Library, eCampus Ontario is used under a CC BY 4.0 license.

Chapter 3

- *Research Methods in Psychology (4th ed.)* by Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler; and Dana C. Leighton (2019), Kwantlen Polytechnic University is used under a CC BY-NC-SA 4.0)
- From the "replicability crisis" to open science practices. In *Research Methods in Psychology – 2nd Canadian Edition* by I-Chant A. Chiang; Rajiv S. Jhangiani; and Paul C. Price is used under a CC BY-NC-SA 4.0 license.
- 1.1 Psychology as a science. In *Psychology – 1st Canadian Edition* by Sally Walters (2020), Thompson Rivers University is used under a CC BY-NC-SA 4.0 license. https://psychology.pressbooks.tru.ca/chapter/1-1-psychology-as-a-science/
- Absolute knowledge. In *Introduction to History and Philosophy of Science* (Barseghyan, Overgaard & Rupik, 2018), Open Library, eCampus Ontario is used under a CC BY 4.0 license.

Chapter 4

- "1.1 Psychology as a Science" in *Psychology – 1st Canadian Edition* (Walters, 2020), via Thompson Rivers University is used under a CC BY-NC-SA 4.0 license.

- *Introduction to Logic and Critical Thinking* by Matthew J. Van Cleave (2016), B.C. Open Collection is used under a CC BY-NC-SA 4.0 license.

Chapter 5

- Chapter "12. *Analyzing the Data*" *in* *Research Methods in Psychology* by Jhangiani et al. (2019), via Kwantlen Polytechnic University is used under a CC BY-NC-SA 4.0 license.
- Chapter "3. Scientific Method" in *Introduction to History and Philosophy of Science* by Barseghyan et al. (2018), via eCampusOntario, is used under a CC BY 4.0 license.

Chapter 6

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.

Chapter 7

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.

Chapter 8

- *Research Methods in Psychology, 4th edition by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.*

Chapter 9

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.

Chapter 10

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019)
- 2.3 Analyzing Findings, In *Psychology 2e* by Rose M. Spielman, William J. Jenkins, & Marilyn D. Lovett (2020), is used under a CC BY 4.0 license.

Chapter 11

- *Principles of Social Psychology (1st International H5P Edition)* by Dr. Rajiv Jhangiani and Dr. Hammond Tarry (2022), BCcampus, is adapted under a CC BY-NC-SA 4.0 license.

If you redistribute all or part of this book, it is recommended the following statement be added to the copyright page so readers can access the original book at no cost:

Download for free at https://criticalthinking.pressbooks.tru.ca/

Sample APA-style citation (7th Edition):

Ramoo, D., (2025). *Critical thinking*. Thompson Rivers University Open Press. https://criticalthinking.pressbooks.tru.ca/

Cover image attribution:

Critical Thinking © Dinesh Ramoo, 2025.

This book was produced with Pressbooks (https://pressbooks.com) and rendered with Prince.

Contents

VI. Descriptive Research

VII. Correlations

VIII. Experiments

IX. Interpreting Evidence

X. Peer Review

XI. Cult Classics

XII. Communicating About Science

Critical Thinking is an Open Educational Resource (OER) textbook focused on Critical Thinking with a particular emphasis on Psychology and the Social Sciences. It addresses the need for students to critically evaluate information in an era where access to information is vast, but misinformation is prevalent. This resource also attempts to introduce students to the universality of critical thinking instead of the narrower views prevalent in most textbooks. The textbook combines new material authored by the project lead, Dinesh Ramoo, with adaptations from other OER resources.

Acknowledgments

The Open Press

The Open Press combines TRU's open platforms and expertise in learning design and open resource development. TRU Open Press supports the creation and reuse of open educational resources, while encouraging open scholarship and research.

Land Acknowledgement

Thompson Rivers University (TRU) campuses are situated on the ancestral lands of the Tk'emlúps te Secwépemc and the T'exelc within Secwepemcúl'ecw, the ancestral and unceded territory of the Secwépemc. The rich tapestry of this land also encompasses the territories of the St'át'imc, Nlaka'pamux, Tŝilhqot'in, Nuxalk, and Dakelh. Recognizing the deep histories and ongoing presence of these Indigenous peoples, we express gratitude for the wisdom held by this land. TRU is dedicated to fostering an inclusive and respectful environment, valuing education as a shared journey. TRU Open Press, inspired by collaborative learning on this land, upholds open principles and accessible education, nurturing respectful, reciprocal relationships through the shared exchange of knowledge across generations and communities.

Resource Development Team 2024

Authors: Dinesh Ramoo, PhD, FHEA

Publishing Manager: Dani Collins, MEd
Copy Editing: Kaitlyn Meyers, BA
Production: Jessica Obando Almache, BCS

Resources

Critical Thinking by Dinesh Ramoo has been created from a combination of original content and materials compiled and adapted from several open-text publications, including:

Chapter 1

- An Introduction to Philosophy, by W. Russ Payne (2015), BC Campus, is used under a CC BY-NC 4.0 license.
- Introduction to History and Philosophy of Science. by Hakob Barseghyan, Nicholas Overgaard, & Gregory Rupik (2018), Open Library, eCampus Ontario, is used under a CC BY 4.0 license.

Chapter 2

- Absolute knowledge. In *Introduction to History and Philosophy of Science* (*Barseghyan, Overgaard & Rupik*, 2018), Open Library, eCampus Ontario is used under a CC BY 4.0 license.

Chapter 3

- *Research Methods in Psychology* (4th ed.) by Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler; and Dana C. Leighton (2019), Kwantlen Polytechnic University is used under a CC BY-NC-SA 4.0)
- From the "replicability crisis" to open science practices. In *Research Methods in Psychology – 2nd Canadian Edition*by I-Chant A. Chiang; Rajiv S. Jhangiani; and Paul C. Price is used under a CC BY-NC-SA 4.0 license.
- 1.1 Psychology as a science. In *Psychology – 1st Canadian Edition* by Sally Walters (2020), Thompson Rivers University is used under a CC BY-NC-SA 4.0 license.

https://psychology.pressbooks.tru.ca/chapter/1-1-psychology-as-a-science/

- Absolute knowledge. In *Introduction to History and Philosophy of Science* (*Barseghyan, Overgaard & Rupik*, 2018), Open Library, eCampus Ontario is used under a CC BY 4.0 license.

Chapter 4

- "1.1 Psychology as a Science" in *Psychology – 1st Canadian Edition* (Walters, 2020), via Thompson Rivers University is used under a CC BY-NC-SA 4.0 license.
- *Introduction to Logic and Critical Thinking* by Matthew J. Van Cleave (2016), B.C. Open Collection is used under a CC BY-NC-SA 4.0 license.

Chapter 5

- Chapter "12. Analyzing the Data" in Research Methods in Psychology by Jhangiani et al. (2019), via Kwantlen Polytechnic University is used under a CC BY-NC-SA 4.0 license.
- Chapter "3. Scientific Method" in Introduction to History and Philosophy of Science by Barseghyan et al. (2018), via eCampusOntario, is used under a CC BY 4.0 license.

Chapter 6

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0

license.

Chapter 7

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.

Chapter 8

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.

Chapter 9

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.

Chapter 10

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019)
- 2.3 Analyzing Findings, In *Psychology 2e* by Rose M. Spielman, William J. Jenkins, & Marilyn D. Lovett (2020), is used under a CC BY 4.0 license.

Chapter 11

- *Principles of Social Psychology (1st International H5P Edition)* by Dr. Rajiv Jhangiani and Dr.

Accessibility

The web version of *Critical Thinking* has been designed to meet Web Content Accessibility Guidelines 2.0, level AA. In addition, it follows all guidelines in Appendix A: Checklist for Accessibility of the Accessibility Toolkit – 2nd Edition.

Includes:

- **Easy navigation.** This resource has a linked table of contents and uses headings in each chapter to make navigation easy.
- **Accessible videos.** All videos in this resource have captions.
- **Accessible images.** All images in this resource that convey information have alternative text. Images that are decorative have empty alternative text.
- **Accessible links.** All links use descriptive link text.

Accessibility Checklist

Element	Requirements	Pass
Headings	Content is organized under headings and subheadings that are used sequentially.	Yes
Images	Images that convey information include alternative text descriptions. These descriptions are provided in the alt text field, in the surrounding text, or linked to as a long description.	Yes
Images	Images and text do not rely on colour to convey information.	Yes
Images	Images that are purely decorative or are already described in the surrounding text contain empty alternative text descriptions. (Descriptive text is unnecessary if the image doesn't convey contextual content information.)	Yes
Tables	Tables include row and/or column headers with the correct scope assigned.	Yes
Tables	Tables include a title or caption.	Yes
Tables	Tables do not have merged or split cells.	Yes
Tables	Tables have adequate cell padding.	Yes
Links	The link text describes the destination of the link.	Yes
Links	Links do not open new windows or tabs. If they do, a textual reference is included in the link text.	Yes
Links	Links to files include the file type in the link text.	Yes
Video	All videos include high-quality (i.e., not machine generated) captions of all speech content and relevant non-speech content.	Yes
Video	All videos with contextual visuals (graphs, charts, etc.) are described audibly in the video.	Yes
H5P	All H5P activities have been tested for accessibility by the H5P team and have passed their testing.	Yes
H5P	All H5P activities that include images, videos, and/or audio content meet the accessibility requirements for those media types.	Yes
Font	Font size is 12 point or higher for body text.	Yes
Font	Font size is 9 point for footnotes or endnotes.	Yes
Font	Font size can be zoomed to 200% in the webbook or eBook formats.	Yes

Mobile Check	Layout displays properly on smaller screen sizes and is mobile-friendly.	

Known Accessibility Issues and Areas for Improvement

- Tables use merged cells but they have been structured to work properly with screen readers – make sure tables do NOT have merged cells!
- The following videos do not have edited captions:

Adapted from the Accessibility Toolkit – 2nd Edition by BCcampus, licensed under CC-BY.

Other Formats Available

- In addition to the web version, this book is available in a number of file formats, including PDF, EPUB (for eReaders), and various editable files. The Digital PDF has passed the Adobe Accessibility Check. Please contact openpress@tru.ca for more details.

I. THE GREAT CONVERSATION

Learning Objectives

- Understand the history of thought and critical thinking
- Define the five characteristics of mythopoeic thought
- Define philosophy and how it differs from other forms of thinking
- Summarise the major developments in the history of philosophy
- Describe how and why philosophy arose in different contexts

Thinking About the World

Most people would rather die than think and many of them do!

— Bertrand Russell, *The ABC of Relativity*

Thinking is hard. As the philosopher Bertrand Russell says, most people would rather not engage in the kind of hard thought that is needed to understand the world. However, this is not a fault but an adaptation. We evolved as human beings to assess the world quickly and make snap judgements. They were the ones who survived in our ancestral past. Those of us who made careful assessments of the situation thoughtfully became lunch. Given this, it is amazing that we also had the ability to become critical thinkers. This gave our ancestors the edge to not just live in the moment but to think in the long-term. They tried to understand the world around them and developed complex societies that allowed them to survive better than other animals.

This book is about a subject that is hard to learn. Most of us think of ourselves as people who make carefully considered decisions. We do not realize how much we are influenced by our culture, family, peers and environment. To get through all these barriers and understand the world as it is can be difficult. But "enter through the narrow gate. For wide is the gate and broad is the road that leads to destruction, and many enter through it" (Matt 7:13). Making the effort to be better thinkers will be its own reward.

As we go through this book, you will encounter ideas and concepts that will help you become better thinkers. At the same time, you will come across ideas that may be uncomfortable and go against what you have been taught in the past. Engage with those ideas. Try to critique them through reason. Make sure you do not create mind blocks to shut them out. It is always possible that you are correct and I am wrong. The main goal is to make sure we arrive at the truth through reason. In other words, you will learn not *what* to think but *how* to think.

Mythopoeic Thought

I do not know where to find in any literature, whether ancient or modern, any adequate account of that nature with which I am acquainted. Mythology comes nearest to it of any.

— Henry David Thoreau, *The Journal of Henry David Thoreau, 1837–1861*

Throughout human history, people have come up with ideas about how the world works. We call these stories myths (somewhat disparagingly). However, myths were ways of understanding the world with the best tools available to our ancestors. How does the sun come up every day? What is the moon? Why do people sometimes behave different? What causes them to hear disembodied voices? Why do we get sick?

Before we delve into the development of philosophy and its successor, natural philosophy (or science), we need to understand what makes science (and philosophy) different from other ways of thinking. Indeed, these other ways of explaining and understanding have been the dominant way of knowing for most of human history. The push towards a more rational and systematic way of thinking does not appear to be a natural instinct for human beings. In fact, the recent increase in pseudoscientific beliefs and conspiracy theories is not new, but a perfectly normal phenomenon. Therefore, we will start by defining what these other ways of knowing are in order to see how philosophy started to move humanity towards rationality and (later) empiricism.

If you are used to hearing people talk, then when you hear voices when no one is around, it is logical to assume that there must be disembodied spirits. Things move because we act upon them. If the sun and the moon are moving, then it is logical to assume that *someone* must be moving them. Such explanations are what we call mythopoeic thought. The term comes from a poem written by J. R. R. Tolkien. Mythopoeic thought involves explanations about the world and the origins of things as told through stories. We will

start by looking at the key characteristics of mythopoeic thought (Frankfort et al., 1946/1977).

1. Myths Are Stories About Persons

Mythic traditions are tales of gods, heroes, mythical creatures, and ordinary people. Things are the way they are because these individuals did things in the mythic past. In Ancient Israel, the rainbow was thought to be a reminder to their god Yahweh to not flood the Earth again when it rains (Genesis 9:16). Some cultures think of the rainbow as the bow of a god. The Arabic term قزح قوس , *Qaws Quzaḥ* still means "the bow of (the god) Quzaḥ". Similarly, the Hindi term ◇◇◇◇◇◇◇◇◇ *indradhanush* means "the bow of (the god) Indra".

In the same vein, natural phenomena are explained through such mythic beings. Among the Algonquian people, thunder is caused by the flapping of the wings of thunderbirds (Cleland et al., 1984) and lightening comes from their eyes (Lenik, 2012).

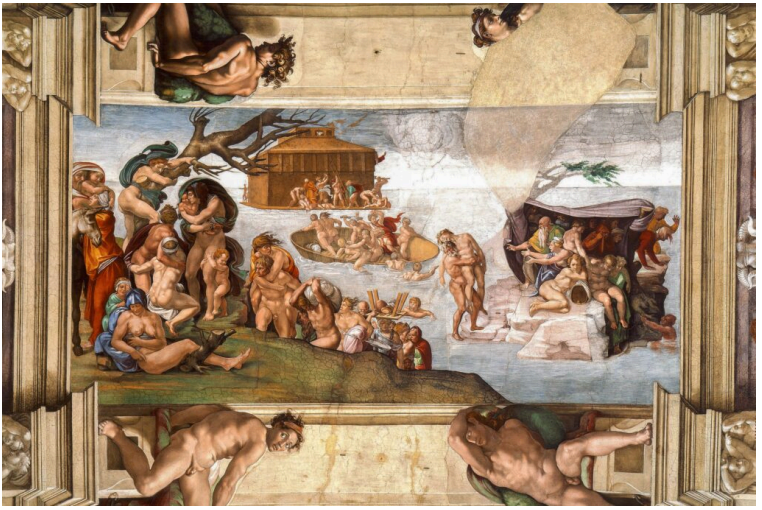


Figure 1.1 *The Deluge in the Noah Story as painted my Michelangelo in the Sistine Chapel. Genesis 6-9 gives two interwoven versions of the flood story. One where Noah takes two of each animal and another where he takes two of each unclean and seven of each clean animal. In one version the flood lasts for 40 days and night while in the other it lasts for 150 days. (Michelangelo/ Wikimedia Commons)* Public Domain

2. Myths Have a Multiplicity of Explanations for the Same Phenomenon

Mythic traditions have multiple stories that are not logically exclusive. Numerous stories may exist in parallel. For example, we have two distinct genealogies for Jesus in the gospels of Matthew and Luke along with very distinct nativities. The Ancient Greeks had numerous stories about the birth of Zeus or about the adventures of Hercules. The Ancient Egyptians considered the sun to be the god *Ra*, a ball pushed through the sky by a giant scarab beetle *Khepri*, and the eye of the hawk *Horus*.

Within traditions, there is rarely any attempt to harmonise these

contradictory stories. They coexist to tell us different versions of the way things were and are in the world around us.

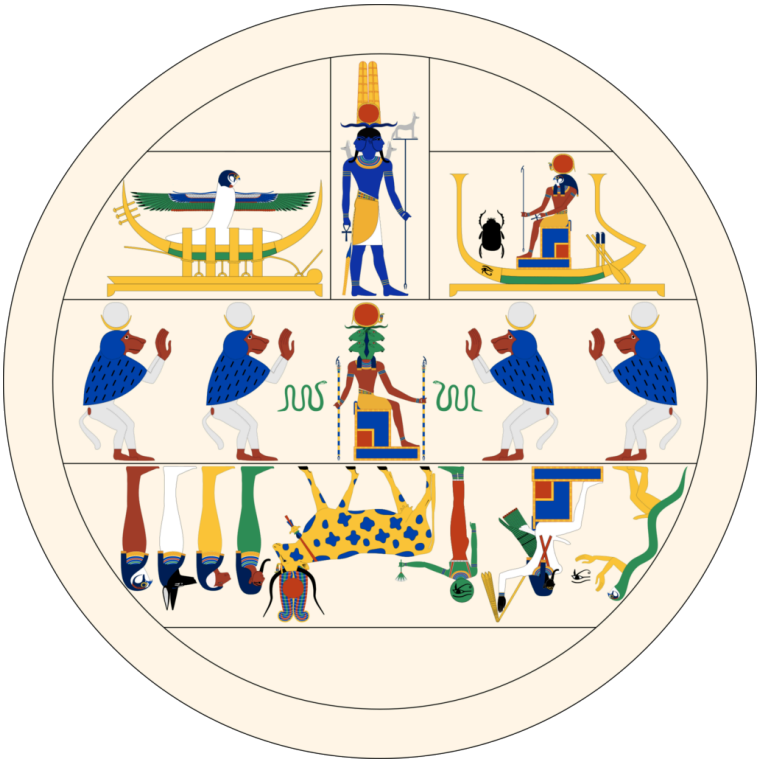


Figure 1.2 *Multiple versions of the sun god in Ancient Egypt. We see Amun (blue), Ra (hawk headed), and Khnum (ram headed), all with the solar disk on their head. Baboons were associated with the sun as they have the habit of sunbathing which looked like worship of the sun to the Ancient Egyptians. All of these contradictory mythical depictions of the sun existed simultaneously in Ancient Egyptian religion. (User:HarJIT/Wikimedia Commons)*

3. Myths Are Conservative to Change

Myths do not change easily. In fact, there is a resistance to change. The stories that were told thousands of years ago may still be told

today within certain communities. This is because these stories continue to give meaning to their respective communities. However, this means that as new ideas emerge, these traditions may act as a barrier to changing people's views about how the world works.

***Figure* 1.3** *Raven and the Emergence of the First people by Bill Reid. This myth told by the Haida people of the Pacific Northwest Coast may have been passed on from generation to generation with little change. (Bill Reid [artist]/Scarlett Sapho [photographer]/ Wikimedia Commons)* CC BY-SA 2.0

4. Myths Are Self-Justifying

The only justification for the truth of a particular tradition may be that it was inspired by the gods. How do we know that Homer knew all the activities of the heroes of the Iliad? He tells us at the very beginning by invoking the muse that inspires him to see what happened (μῆνιν ἄειδε θεὰ, "Sing of Anger, O Goddess").

Figure 1.4 *An Angel piercing the Heart of St. Teresa of Avila with Divine Inspiration by Bernini. The statements of seers and prophets were to be believed because they were thought to be inspired by divine beings. (Gian*

Lorenzo Bernini (1598–1680) [artist]/Livioandronico2013 [photographer]/Wikimedia Commons) CC BY-SA 4.0

5. Myths Are Morally Ambivalent

The gods and heroes in mythic traditions do not always do what is moral or admirable. This is true even for the people within those mythic traditions. In Homer's *Iliad*, Achilles is the hero, but his actions are far less admirable even to the Greeks. In the traditions of the Ancient Israelites, Yahweh commands wholesale genocide of the Canaanites as well as the sexual exploitation of their women (Numbers 31.18). Such examples abound in any myth you will come across (unless it was scrubbed clean for general consumption by modern writers).

Figure 1.5 *The Binding of Isaac by Caravaggio. Yahweh asks for Abraham's son as a burnt sacrifice. There is no condemnation of this in the narrative and Abraham is lauded for his devotion. (Caravaggio (1571–1610)/Wikimedia Commons)* CC BY-SA 3.0

This list is not an attempt to denigrate mythic traditions. Rather, it is a list of common characteristics that all mythopoeic thought shares regardless of culture or community. They are reflective of human nature as it is and honest about the brutality of life. But at some point in history, three civilizations started a process that led to what we now call *philosophy*. A tradition that was very different from what came before it.

Myth Quest

Think of a myth you have personally encountered. List how many of the characteristics of mythopoeic thought you can find in it. Also, try to find other versions of that myth and see why they would have been changed to suit the needs of the community at different times.

Media Attributions

- **Figure 1.1** The Deluge after restoration [Part of Sistine Chapel ceiling] by Michelangelo (1475–1564) [Photo from the Web Gallery of Art], via Wikimedia Commons, is in the public domain.
- **Figure 1.2** Egyptian Hypocephalus Images by User:HarJIT [Derivative of files from Jeff Dahl, A worried citizen, RootOfAllLight, Finn Bjørklid, FDRMRZUSA, Jasmina El Bouamraoui and Karabo Poppy Moletsane, and Rawpixel], via Wikimedia Commons, is used under a CC BY-SA 4.0 license.
- **Figure 1.3** The Raven and the First Men, Museum of

Anthropology (7960613690) by Bill Reid [Photo by Scarlett Sapho], via Wikimedia Commons, is used under a CC BY-SA 2.0 license.

- **Figure 1.4** Ecstasy of St. Teresa HDR by Gian Lorenzo Bernini (1598–1680) [Photo by Livioandronico2013], via Wikimedia Commons, is used under a CC BY-SA 4.0 license.
- **Figure 1.5** The Sacrifice of Isaac by Caravaggio by Caravaggio (1571–1610) [Photo from Nicolas Pioch/WebMuseum/ibiblio.org], via Wikimedia Commons, is in the public domain.

The First Flowerings of Philosophy

The First Flowerings of Philosophy

What seest thou else
In the dark backward and abysm of time?
— William Shakespeare, *The Tempest*

What is philosophy? You might have heard people say, "Never be late, that's my philosophy." That is a generous term for such a banal statement. There must be something more to it. Literally, the term *philosophy* means *love of wisdom* from the Ancient Greek φίλος (philos: 'love') and σοφία (sophia: 'wisdom'). Is that it? What makes philosophy different from what came before it? Does every culture inevitably come up with philosophy or is it a phenomenon that arose in particular contexts?

As far as we know, only three civilizations have come up with philosophical traditions. What we now call philosophy and science (what used to be called natural philosophy) are an amalgamation of ideas and concepts that came from these three traditions: Ancient Greeks, Indians, and Chinese. What differentiated these philosophies from what came before them? Let us explore our criteria for mythopoeic thought again and see where philosophy differed:

1. Myths are stories about persons, whereas **philosophy tries to explain natural phenomena in natural terms:** Anaximander of Miletus (c. 550 BCE) suggested that the Earth was a finite body with the sky above and below it. This is an extreme leap of imagination given that most people thought of the Earth an expanse with water above and below (Genesis 1:7). Instead of

making the sun and moon into gods, Anaxagoras (c. 500–428 BCE) stated that the sun was made of molten metal, the moon was a rock, and the stars were hot stones.

2. Myths have a multiplicity of explanations for the same phenomenon, while **philosophy tries to create systematic and internally consistent explanations:** Within each philosophical tradition, there is an attempt to create systematic and logical ideas that do not contain contradictions. In fact, internal contradictions were one of the main criteria for rejecting a postulate.
3. Myths are conservative to change, whereas **philosophy is all about new ideas developing all the time:** Within the first century of Greek philosophy, Thales, Anaximander, Anaximenes, Xenophanes, Pythagoras, and Heraclitus had come up with different explanations about the origins of the universe. Numerous schools of thought (each challenging the other) had emerged in India within the lifetime of the Buddha (along with before and after him). During the late spring and autumn period of Ancient China, there were the ◇◇◇◇ *zhūzǐ bǎijiā* or hundred schools of thought which predated Confucianism and Mohism.
4. Myths are self-justifying, while **philosophy looks to give rational augments:** While early philosophical texts—such as Parmenides of Elea's Way of Truth, The Analects of Confucius and the Māṇḍūkya Upanishad—are aphoristic statements, later philosophical texts are characterised by substantial time spent reasoning through arguments to reach a conclusion. The idea is that you do not believe something because of the reputation or status of the speaker (or writer) but because of the reasons and evidence they present.
5. Myths are morally ambivalent, whereas **philosophy spends a lot of time trying to define what is good:** The Ancient Greek, Indian, and Chinese philosophers spend a lot of time criticizing immorality in their mythology. Socrates refuses to believe that the shameful acts ascribed to the Greek gods could be true and

spends considerable time on defining *The Good* and how to live the good life. Indian philosophy is filled with discussion on what is *dharma*. Chinese philosophers discuss endlessly about ◈*ren*, 'humaneness,' and ◈/◈*yi*, righteousness.

Myth and Philosophy

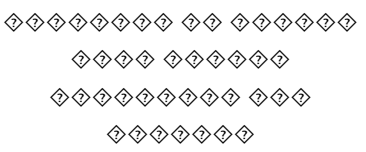
Before going any further, try to remember what you thought of as philosophy before reading this chapter. Write down what you now think has changed your perception of this definition.

Now, see if you can think of myths that you grew up hearing or seeing on television or in movies. Use the five points listed above to analyse them and see why they are different from philosophy.

As we can see from the five characteristics that changed with philosophy, there appears to be an underlying thread that connects different philosophy. This would be the application of critical thinking to pre-existing ideas. Critical thinking could be defined as the application of rational, sceptical, and unbiased analyses and evaluation in order to make judgements about the world. This involves the analysis of facts, evidence, observations, and arguments. Critical thinking requires effective communication and systematic ways of thinking as well as a commitment to overcoming biases.

To say that philosophy only arose among the Ancient Greeks, Indians and Chinese does not mean that other cultures and communities are inferior to them in some way. Rather, it is to realize that critical thinking and philosophy are often the products of

specific sociocultural influences that need to be nurtured. One reason these civilizations came up with such ideas is that they were multicultural and open to new ideas. These civilizations traded outside of their immediate cultural spheres and interacted with other cultures and communities. Greek traders settled in Asia Minor (what is now Türkiye) and along the Northern Coast of Africa. They talked to Egyptians and Ethiopians (a general term for the people of sub-Saharan Africa) the latter being referred to as the most pious people on earth by Homer. Greeks and Indians exchanged ideas on mathematics for centuries. Mathematical treatises were translated from Ancient Greek into Sanskrit. Indians travelled to China and vice versa. They not only talked to each other, but appreciated each other's contribution as seen here:



The Yavanas (Greeks) are, no doubt, barbarians. However, their contribution to this science (of mathematics and astronomy) is incontrovertible.

— Varahamirira, *Brahasamhita* (2:32)

This exchange of ideas and appreciations for concepts from other cultures was the driving force for philosophy to emerge. Whenever such openness to other cultures and ideas disappeared, then there was stagnation. In the same vein, just because a clear philosophical tradition is absent to begin with in some cultures does not mean that there was no critical thinking or attempts to go beyond tradition. We can point to the wisdom literature in the Bible as an example of such an attempt. But they were always subsumed by traditional conservatism against challenging the status quo.

We also see rational discussion as a driving force for Indigenous communities in North America. Early missionaries were surprised at how adept Indigenous leaders were in challenging the arguments they offered in favour of converting to Christianity (along with the subsequent acceptance of European religious authority). Reports suggest that Indigenous leaders were able to discuss these issues

rationally as they themselves had gained their leadership positions by demonstrating such abilities within their communities (Graeber & Wengrow, 2023). The only reason we did not see such skills develop into philosophy is because there was not an attempt to codify ideas into schools of thought or an impetus to challenge pre-existing ways of thinking.

Greek Thought

ἔοικα γοῦν τούτου γε σμικρῷ τινι αὐτῷ τούτῳ σοφώτερος εἶναι, ὅτι ἃ μὴ οἶδα οὐδὲ οἴομαι εἰδέναι.

I realized that I was wiser in this sense: that I do not know anything nor do I think that I do.

— Socrates, *Plato's Apology* (2d)

The Greeks and, in particular, the Athenians are credited with the flowering of philosophy in the West. Part of this had to do with the environmental conditions of Greece: an arid land which bore mostly olives and vines. However, these could be converted into oil and wine to be traded for other goods across the Mediterranean and later across the ocean towards India. They created colonies throughout Asia Minor, Italy, and the northern coastline of Africa. In fact, the first Greek natural philosopher (or as we would understand it: scientist), Thales, lived not in Greece but in Asia Minor (modern day Türkiye). The interactions with other cultures challenged the prevailing norms of the day and numerous philosophers arose throughout the Hellenistic world. The most influential of these was Socrates.

Socrates asked questions. In fact, he was famous for putting down the proud and boastful Athenian gentry with his wit. However, he was not humiliating them for no reason; instead, he was inquiring as to whether they were as wise as they claimed. He developed the *Socratic method*, which is a form of argumentative dialogue between individuals. It searches for commonly held truths and scrutinizes them for consistency. Through such refutations and logical analyses, it is assumed that a truth will emerge. We do not know anything about Socrates with any certainty as he never wrote anything down. However, he influenced many Athenians, including a youth named Plato. But not everyone was impressed, and eventually, some aristocratic youths brought a lawsuit against him for blaspheming against the gods of the city and corrupting the youth. Instead of pleading for his life in front of his fellow citizens (as was often the case back then), Socrates gave a resounding defence

(*apologia*) of his life and philosophical inquiries. He was condemned to death by drinking hemlock. His friends tried to bribe the guards and smuggle him away into exile, but Socrates refused and bravely met his undeserved end by drinking the poison.

Figure 1.6 *The Death of Socrates (French: La Mort de Socrate) painted by French painter Jacques-Louis David. This scene, where the great philosopher bravely meets his death while defending his ideas to the bitter end, has been as significant to thinkers throughout the ages as the crucifixion is to Christians. (Jacques-Louis David (1748–1825)/Wikimedia Commons)* Public Domain

Understanding the Socratic Method

The best way to understand Socrates and his method is to look at an example. One of the earliest dialogues of Socrates is *Euthyphro*. Socrates meets a young man,

Euthyphro, outside the public courthouse and engages him in conversation about piety. The dialogue goes as follows (the numbers and letters are Stephanus pagination which is system for referencing translations of Plato):

Ironic Pose: "I must become your pupil." (5a-5d)

Euthyphro claims to know all about the subject, and Socrates states his desire to learn from him.

First Definition: "The Pious is what I am doing." (5d-6d)

Euthyphro is prosecuting his father for murder and claims that to do so even to one's own family must be piety. He gives as examples, such as the fact that the supreme god Zeus punished his father Cronus for his wickedness. However, Socrates refuses to believe that such myths are true.

Second Definition: "The Pious is what is dear to the gods." (6c-8b)

Euthyphro now modifies the definition as above. However, Socrates asks how we can be sure of what is dear to the gods given that there are lots of gods and they might disagree.

Correction: "The gods don't disagree about penalizing injustice." (8b-9a)

Euthyphro claims that all the gods would agree that murder should be punished.

Question: "What do the gods agree on in the case?" (9a-9b)

Socrates agrees that no one refutes that wrongdoers must be punished. The issue is that the other party will claim not to have done any wrong at

all. So, what part of this case would all the gods agree on?

Third Definition: "The Pious is what is loved by all the gods." (9c-10a)

Euthyphro now claims that the pious is what is loved by the gods.

Question: "Is the Pious loved because it is Pious, or Pious because it is loved?" (10a-11b)

Now, we come to the heart of the matter: is something good because it is loved by the gods or is it loved by the gods because it is good? In other words, is there some inherent property of goodness that is merely recognized by the gods or can anything be good simply because it is approved by the gods? If the former is true, then we can try to define goodness independent of gods.

Suggestion: "Piety is a kind of justice." (11e-12e)

Fourth Definition: "The Pious is the part of justice concerned with the care of the gods." (12e-13d)

Euthyphro claims that piety is a kind of justice. Socrates asks what part of justice constitutes piety.

Fifth Definition: "The Pious is the part of justice concerned with the service of the gods." (13d-14a)

Euthyphro claims that piety is that part of justice that is in service to the gods. Socrates asks what kind of service the gods require.

Sixth Definition: "Piety is knowledge of how to and sacrifice and pray." (14a-15b)

Euthyphro now claims piety is about knowing how

to sacrifice and pray. Socrates asks whether such sacrifice and prayer are required by the gods in any way. In other words, is it a kind of business where we give something to the gods and get something in return?

Seventh Definition: "Piety is what is dear to the gods." (15b-15c)

Euthyphro goes back to his earlier assertion.

Conclusion (15c-16a): "So we must investigate again from the beginning…."

Socrates points out that they are where they began. Euthyphro makes a run for it with excuses, while Socrates laments that he was unable to learn from him.

Plato (429–347 BCE) came from a family of high status in ancient Athens. He was deeply moved by the life and death of Socrates and dedicated his life to immortalising his teacher. Some of his early dialogues chronicle events in Socrates' life. Socrates is a character in all of Plato's dialogues. But in many, the figure of Socrates is employed as a voice for Plato's own views. Unlike Socrates, Plato offers very developed and carefully reasoned views about a great many things. Plato's metaphysics and epistemology are best summarized by his device of the divided line (Table 1.1). The vertical line between the columns below distinguishes reality and knowledge. It is divided into levels that identify what in reality corresponds with specific modes of thought.

Table 1.1 Plato's Divided Line

Objects	Modes of Thought
The Forms	Knowledge
Mathematical objects	Thinking
Particular things	Belief /Opinion
Images	Imaging

Here, we have a hierarchy of Modes of Thought, or types of mental representational states, with the highest being knowledge of the forms and the lowest being imaging (in the literal sense of forming images in the mind). Corresponding to these degrees of knowledge we have degrees of reality. The "less real" includes the physical world, and "even less real", our representations of it in art. The "more real" we encounter as we inquire into the universal natures of the various kinds of things and processes. According to Plato, the only objects of knowledge are the forms, which are abstract entities.

In saying that the forms are abstract, we are saying that while they do exist, they do not exist in space and time. They are ideals in the sense that a form, say the form of horse-ness, is the template or paradigm of being a horse. All the physical horses partake of the form of horse-ness but exemplify it only to partial and varying degrees of perfection. No actual triangular object is perfectly triangular, for instance. But all actual triangles have something in common, triangularity. The form of triangularity is free from all of the imperfections of the various actual instances of being triangular. We get the idea of something being more or less perfectly triangular. For various triangles to come closer to perfection than others suggests that there is some ideal standard of "perfectly triangularity." For Plato, this is the form of triangularity. Plato also takes moral standards, like justice, and aesthetic standards, like beauty, to admit of such degrees of perfection. Beautiful physical things all partake of the form of beauty to some degree or another.

But all are imperfect in varying degrees and ways. Knowledge of the nature of the forms is a grasp of the universal essential natures of things. It is the intellectual perception of what various things, like horses or people, have in common that makes them things of a kind. Plato accepts Socrates' view that to know the good is to do the good. So, his notion of epistemic excellence in seeking knowledge of the forms will be a central component of his conception of moral virtue.

Plato's most illustrious student, Aristotle, is a towering figure in the history of philosophy and science. Aristotle made substantive contributions to just about every philosophical and scientific issue known in the Ancient Greek world. Aristotle was the first to develop a formal system of logic. The study of critical thinking is dependent on understanding logical argumentation. Logic involves analysing arguments and deciding on their correctness. Another definition of critical thinking could be *logically correct thinking* (Wisdom, 2015).

As the son of a physician, Aristotle pursued a life-long interest in biology. His physics was the standard view through Europe's Middle Ages. Though he was a student of Plato, he rejected Plato's other-worldly theory of forms in favour of the view that things are a composite of substance and form. Contemporary discussions of the good life still routinely take Aristotle's ethics as their starting point. Below is a brief sketch of Aristotle's logic:

Aristotle's Logic

Authoritative for well over 2,000 years, the core of Aristotle's logic is the systematic treatment of categorical syllogisms. Consider this argument:

1. All monkeys are primates.
2. All primates are mammals.

3. So, all monkeys are mammals.

This argument is a categorical syllogism. That is a rather antiquated way of saying it is a two-premise argument that uses simple categorical claims. Simple categorical claims come in one of the following four forms:

1. All A are B
2. All A are not B
3. Some A are B
4. Some A are not B

There are a limited number of two premise argument forms that can be generated from combinations of claims having one of these four forms. Aristotle systematically identified all of them, offered proofs of the valid ones, and demonstrated the invalidity of the others.

Aristotle's system of logic was not only the first developed in the West, it was considered complete. Beyond this, Aristotle proves a number of interesting things about his system of syllogistic logic, and he offers an analysis of syllogisms involving claims about what is necessarily the case as well. No less an authority than Immanuel Kant, one of the most brilliant philosophers of the 18th century, pronounced Aristotle's logic complete and final. It is only within the past century or so that logic has developed substantially beyond Aristotle's.

While we have mostly talked about Plato and Aristotle, Greek philosophical traditions evolved over the centuries, giving birth to numerous traditions such as the cynics, stoics and Neo-Platonists. They spread through Alexander the Great's empire and went on to be a part of the Roman world, eventually influencing middle-eastern religions such as Judaism, Christianity, and Islam. These

ideas even spread as far east as India, where they met another ancient civilization with its own philosophical tradition.

Media Attributions

- **Figure 1.6** David – The Death of Socrates by Jacques-Louis David (1748–1825) [Photo from the Metropolitan Museum of Art], via Wikimedia Commons, is in the public domain.

Indian Thought

But then, who really knows and who can here proclaim it? Whence comes this world and how was it created?

Surely the gods came after creation.

So, who really know whence it arose?

Where did creation originate? Maybe there is a creator who fashioned it, and maybe he did not.

The ultimate creator who sees from the highest heaven, maybe he knows

– or maybe he knows now.

— Nasadiya Sukta, *Rigveda* (10:129)

The above verse from the Hindu sacred text Rig Veda shows the early blossoms of critical inquiry on the subcontinent. The author is not simply describing the creation of the universe as a story but actively questioning the very idea of whether it was created or not. Not only that, he is even questioning whether a creator exists or whether that creator knows about creation. This kind of questioning and challenging of tradition led to a cascade of philosophical texts around 500 BCE. Numerous schools of thought emerged with their own ideas about reality and how to understand it. Some of these were absorbed into mainstream religion, which evolved into what we now call Hinduism. Overall, there were six schools of orthodox or theistic inquiry and several heterodox schools that rivalled them (Doniger, 2014; Perrett, 2000).

Orthodox/Theistic Schools of Thought

Table 1.2 Orthodox/Theistic Schools of Thought

School of Thought	Important Scholar(s)	Tenants
Nyaya	Gautama, Udayana	The Indian school of logic.
Vaisheshika	Kaṇāda, Candra	Examined the universe in terms its smallest constituents: atoms. Accepted only direct observation and inference as the means of knowledge.
Samkhya	Kapila, Pancasikha	Postulated that all matter was made up of three gunas or qualities: sattva, rajas, and tamas
Yoga	Patanjali	The universe was thought to be dualistic: made of *purusha* (consciousness) and *pratkrti* (matter). Salvation could be achieved through meditative practices.
Mīmāmsa	Jaimini	Based on interpreting the meaning of sacred scriptures and rituals.
Vedanta	Baudhayana, Adi Sankara, Ramanuja, Madhva	A collection of different philosophies that postulates a supreme consciousness as the ultimate reality. The ability to know and experience it is thought be salvation.

These six schools of philosophy continued to engage with each other for two millennia as they evolved different ideas about reality. Nyaya, or logic, became an important component in all schools of thought. Indian ideas about knowledge accepted six means of knowledge (□□□□□□, Pramāṇa). These were Pratyakṣa (perception), Anumāṇa (inference), Upamāṇa (analogy), Śabda (authority or the testimony of experts), Arthāpatti (postulation from circumstances), and Anupalabdhi (non-perception). Of these, direct perception was thought to be both external (from the senses) and internal (cognition). For example, seeing something burning is direct observation. Inference would be an indirect method of knowledge (e.g., seeing smoke and inferring that something is burning). Upamana, or analogy, would be where an example is given as a

comparison (e.g., A zebra is an animal that looks almost like a cross between a horse and a donkey). Śabda means reliable authority, which could include accepted scripture or experts. Arthapatti, or postulation, means that you infer something from the context. For example, we can calculate the occurrence of eclipses based on the circumstances of where the sun, moon and Earth were during past eclipses. Anupalabdhi, or non-perception, means deriving proof through the non-perception of a thing (e.g., I do not see fire, so there must be nothing on fire). Of these six, the latter two were sometimes rejected by some scholars as inefficient means of knowledge (Flood, 1996).

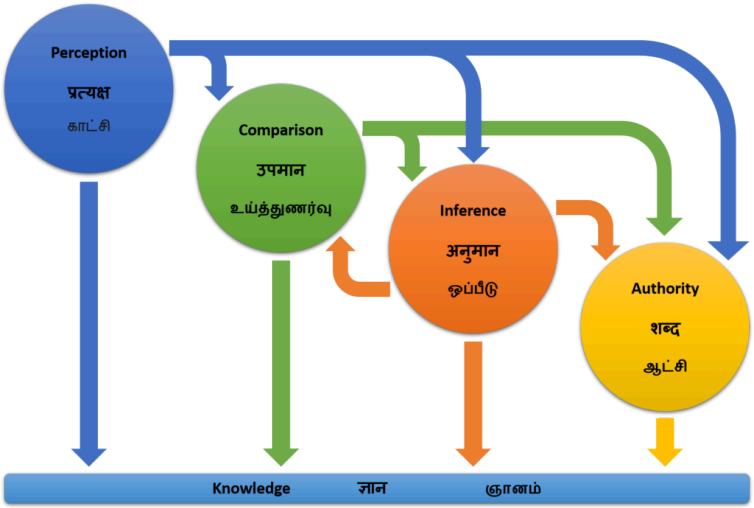


Figure 1.7 *The Nyaya (or logic) school considers perception, inference, comparison/analogy, and revelatory scriptures as four means to correct knowledge (by author).*

Indian philosophers used numerous means of understanding the universe. For example, consider the following argument from the Vaisheshika school.

The Nature of Matter

If you keep cutting something into smaller and smaller pieces, could you keep on doing so continuously or is there a limit beyond which you could not go? This cannot be done through experiments as we cannot see beyond a certain limit.

Assume that matter is continuous. In this case, if you keep cutting a stone into smaller and smaller particles, you would end up with an infinite number of particles (since matter is continuous). Now, the Himalayan Mountain range could also be cut into smaller and smaller particles and you would end up with an infinite number. So, logically, you can start with a stone and end up with the same number of particles as the Himalayas which is ridiculous. Therefore, the original assumption that matter is continuous must be wrong and matter must be made up of a finite number of *paramāṇus* (atoms).

The atoms (or *paramāṇus*) in this case were not the atoms as we know them now. Indian philosophers thought these atoms were for the four elements: earth, air, fire, and water. However, we see that they were deriving these conclusions from logical thinking and not scriptural authority.

Indian schools of logic developed a strong tradition of critical analysis in debating various issues. Indian jurisprudence, for example, used the concept of *purva paksha* (former argument), referred to as the complaint, with other parts of a trial consisting of the *pttara paksha* (the response), *kriyā* (trial or investigation), and *nirnaya* (decision). The same was true of critical arguments in

Indian texts where the *purva paksha* shows that the author has a deep familiarity with the thesis that he or she is criticizing. Only then are you allowed to state your own position vis-à-vis the former thesis. Philosophers who exemplified this technique include Adi Shankara (sixth century CE), Udyotakara (Nyāyavārttika, sixth to seventh century), Vācaspati Miśra (Tatparyatika, ninth century), Ramanujacharya (nonth century), Udayanacharya (Tātparyaparishuddhi, 10th century), Jayanta Bhatta (Nyāyamanjari, ninth century), Madhvacharya (13th century), Visvanatha (Nyāyasūtravṛtti, 17th century), Rādhāmohana Gosvāmī (Nyāyasūtravivarana, 18th century), and Kumaran Asan (1873–1924).

Heterodox Schools of Thought

In addition to the six orthodox schools of thought, there were numerous schools that challenged them. These were often called *Nastika* or non-theistic schools.

Table 1.3 Heterodox Schools of Thought

School of thought	Important scholar(s)	Tenants
Buddhism	Gautama Buddha	Existence is suffering and one can escape the cycle of rebirth by practicing the eightfold path.
Jainism	Mahavira	Similar to Buddhism but advocates extreme nonviolence and austerities as a means of salvation.
Ājīvika	Makkhali Gosala	Postulates that each life is like a pre-woven thread that unfolds as fated.
Charvaka / Lokayata	Brihaspati, Ajita Kesakambalī	Atheism or materialism. Stated that with death all is annihilated so one must live happily.
Ajñana	Sañjaya Belaṭṭhiputta	Agnosticism
Akiriyāvāda	Pūraṇa Kassapa	Denied that there was any good or evil or that there were rewards or punishments for actions.
Sassatavāda	Pakudha Kaccāyana	Matter and souls are eternal and do not interact.

The above table is a highly simplified description of rather complex ideas. However, it gives you an understanding of the diversity of thought. They challenged traditions about the afterlife and the existence of gods. Take for example the following verse:

◻◻-◻◻◻◻◻◻ ◻◻◻
◻◻◻◻◻◻
◻◻◻◻◻◻◻◻◻◻◻◻◻ ◻◻◻◻
◻ ◻◻
◻◻◻-◻◻◻◻◻◻◻◻ ◻◻◻◻◻
◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻◻
◻◻◻◻◻◻◻◻◻

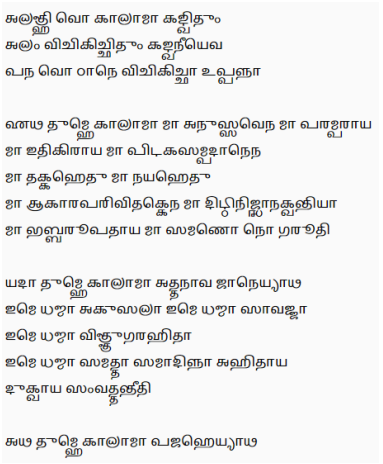
There is no world other than this;
There is no heaven and no hell;

— Sarva-siddhanta Sangraha (2:8)

This statement by atheists in Ancient India would have been unthinkable in most of the ancient world. Indeed, it could get people killed in some parts of the world today! Even the teachers of these schools

The realm of Shiva and like regions, are fabricated by stupid imposters.

exhorted the need for critical thinking as can be seen in the following statement by the Buddha:



— The Buddha, *The Kalama Sutta*
(Translated from Pali by Thanissaro Bhikkhu)

So, in this case, Kalamas, don't go by reports, by legends, by traditions, by scripture, by logical conjecture, by inference, by analogies, by agreement through pondering views, by probability, or by the thought, 'This contemplative is our teacher.'

When you know for yourselves that, 'These qualities are unskilful; these qualities are blameworthy; these qualities are criticized by the wise; these qualities, when adopted & carried out, lead to harm & to suffering' — then you should abandon them.

Such reasoning and the advocacy for it was the hallmark of critical thinking in India and continued for centuries. It allowed philosophers to explore new ideas. For example, Aryabhata (a mathematician from 476–550 CE) challenged the idea of a flat earth and calculated its circumference. His ideas were extrapolated upon by later mathematicians such as Bhaskara I (c. 600 CE) and Nilakantha Somayaji (1465 CE).

Media Attributions

- **Figure 1.7** Created by the author

Chinese Thought

◇◇◇◇◇◇◇◇◇
◇◇◇◇◇ ◇◇◇◇◇◇◇◇◇◇
◇◇◇◇◇◇◇◇◇◇
◇◇◇◇◇◇◇◇

The Master was very ill, and Tzu-lu said that he would pray for him. Confucius said, "was such a thing ever done?" Tzu-lu said, "There is. The Eulogies say: 'I pray for you to the spirits of the upper and lower realm.'" Confucius said, "Then I have been praying for a long time already."

— *Analects* 7:34

The above quote might be considered the essence of Chinese philosophy. Confucius (551–479 BCE) is seen here stating that living a good life is itself enough to serve as prayer and ritual. What we see as Chinese philosophy was the culmination of a large number of schools that emerged during the Warring States era, including Confucianism, Legalism, Agriculturalism, Mohism, Chinese Naturalism, School of Names (Logic), and Taoism (Liang-Hung & Yu-Ling, 2009). This became known as the Hundred Schools of Thought (◇◇◇◇; zhūzǐ bǎijiā). These schools ranged from utopian proto-communalism (Agriculturalism) and being attuned to nature (Taoism) to support for government and state power (Confucianism). In many ways, Confucianism is responsible for the first meritocracy where status was seen as something that could be achieved through hard work and education rather than through birth into the aristocracy (Kung Fu Tze, 1998). Most of the other schools of thought were supressed by various Chinese dynasties and only Confucianism, Taoism, and later Buddhism became ascendent in Chinese society.

The central concept of all Chinese philosophy has been *Tao* or the correct way to live. This was interpreted by Confucius as living within regimented society, while Taoism saw this as an abstract concept lived through non-action (*wu wei*). The Indian concepts from Buddhism—which arrived in China during the Han Dynasty (206 BCE–220 CE)—were added to this. This led to the translation of

numerous Sanskrit texts into Classical Chinese, which then went on to influence Japanese and Korean culture.

Figure 1.8 *The Vinegar Tasters (□□□) painted by a Japanese artist of the Kanō school (16th century). It shows three men, sometimes thought to be Confucious, Laozi and the Buddha tasting a vat of vinegar. Confucius thinks it tastes sour (as Confucianism saw life as sour and in need of rules), the Buddha thinks it tastes bitter (as Buddhism saw life as full of suffering due to material desires), while Laozi tastes the vinegar as vinegar (as Taoism saw life as inherently pleasant if lived according to nature). (Kanō school artist/ Wikimedia Commons)* Public Domain.

Chinese philosophy and natural science continued to develop over the next two millennia. We can thank the Chinese philosophers for inventing gun powder, paper, printing, and the compass (called the Four Great Inventions, □□□□). These inventions have perhaps had the greatest influence on world history. The introduction of paper into the Islamic world led to the writing down of the steps of mathematical proofs (Greek and Indian mathematicians usually gave the final proofs and not necessarily the steps they followed). Later on, the printing press and movable type would allow knowledge to spread among the masses and lead to even greater leaps in technology.

The Three Roots of Modern Philosophy

Now that you have learned about the three civilizations that gave birth to philosophy, see if you can reflect on what has been discussed and come up with some of the differences between them. Consider the following points in your reflection:

- The belief or lack thereof in the supernatural
- The emphasis on respecting ancient traditions

- The impetus to challenge pre-existing beliefs

Media Attributions

- **Figure 1.8** The Three Vinegar Tasters (cropped) by a Kanō school artist during the Mromachi period [Photo from the Tokyo National Musuem], via Wikimedia Commons, is in the public domain.

Medieval Thinking

…ratio in homine sicut Deus in mundo.
…reason is to man what God is to the world.

— Thomas Aquinas in *De Regno* (On Kingship)

The rise and fall of Rome follows the golden age of Ancient Greece. Greek philosophical traditions underwent assorted transformations during this period, but Rome is not known for making significant original contributions to either philosophy or science. Intellectual progress requires a degree of liberty that was not as available in the Roman Empire. Additionally, the intellectual talent and energy available in ancient Rome would have been almost fully occupied with the demands of expanding and sustaining political power and order. Rome had more use for engineers and bureaucrats than scientists and philosophers.

Christianity became the dominate religion in Rome after emperor Constantine converted in the 4th century CE. Also in the 4th century, the great Christian philosopher Augustine, under the influence of Plato, formulated much of what would become orthodox Catholic doctrine. After a rather dissolute and free-wheeling youth, Augustine studied Plato and found much to make Christianity reasonable in it. With the rise of the Catholic Church, learning and inquiry were pursued almost exclusively in the service of religion for well over a millennium. Philosophy in this period is often described as the handmaiden of theology. The relationship between philosophy and theology is perhaps a bit more ambiguous, though. As we have just noted in the case of Augustine, much of Ancient Greek philosophy gets infused into Catholic orthodoxy. But at the same time, the new faith of Christianity spearheads an anti-intellectual movement in which libraries are destroyed and most Ancient Greek thought is lost to the world forever.

Through the West's period of Catholic orthodoxy, most of what we know of Greek science and philosophy, most notably Aristotle's thought, survived in the Islamic world. Islamic philosophy flourished

from the 9th century onwards, with Al-Kindi, until the 12th century CE, with Ibn-Rushd. This golden age of Islamic philosophy (or *falsafa*) was precipitated by the spread of Islam and brought about conditions that fostered new ideas. There was now a multicultural empire that stretched from Iberia (modern day Portugal and Spain) and North Africa to the western parts of Ancient India (what is now Afghanistan). People now had a common *lingua franca*: Arabic. Greek philosophical texts (including Plato and Aristotle) as well as Indian Sanskrit books on mathematics were translated into Arabic and studied by scholars. This is how Indian mathematics and the Indian number system (modern numerals) came to Europe.

The immense contribution of Islamic scholars to the preservation of Ancient Greek and Indian thought cannot be overstated. Eventually, Arabic texts were translated into Latin and studied by Christian monks and scholars. This brought Aristotle's ideas back into the West, while at the same time, they withered in the Islamic worlds as orthodoxy set in. Al-Ghazali's work *Tahafut al-Falasifa* (The Incoherence of the Philosophers) criticized philosophy and directed people towards mysticism and religious orthodoxy. Ibn-Rushd's books were burned by his fellow Muslims and the Islamic world plunged into a dark age from which some parts have yet to emerge.

At the same time, Aristotle's views about the natural world quickly came to be received as the established truth in the Christian world. Aristotle's physics, for instance, became the standard scientific view about the natural world in Europe. Aristotle also wrote about the methods of science and was much more empirical than his teacher, Plato. Aristotle thought the way to learn about the natural world was to make careful observations and infer general principles from them. For instance, as an early biologist, Aristotle dissected hundreds of species of animals to learn about anatomy and physiology.

Scholasticism was a medieval school of philosophy that emerged within the monastic schools. They translated Judaic and Islamic texts into Latin and, in some sense, rediscovered Ancient Greek

thought. Their aim was not understanding the universe per se, but rather to harmonise Aristotle's metaphysical ideas with Christian theology. The Scholastics who studied Aristotle obviously did not adopt the methods Aristotle recommended. The scholastics emphasised dialectical reasoning to extend knowledge by inference and resolve contradictions rather than rely on empirical observations. But some other people disagreed. Galileo, Leonardo da Vinci, and Copernicus were among the few brave souls to turn a critical eye to the natural world itself and, employing methods Aristotle would have approved of, began to challenge the views of Aristotle that the Scholastics had made a matter of doctrine. Thus begins the Scientific Revolution.

Figure 1.9 *Triumph of Saint Thomas Aquinas by Lipo Memmi, circa 1340*
. We see St. Thomas Aquinas in the middle flanked on either side by Plato and Aristotle. We also see the four evangelists (Matthew, Mark, Luke and John with their symbolic animals) as well as Moses (with his ten commandments) and St. Paul above him. On the ground we see the philosopher Averroes (Ibn Rushd) whose work on Aristotle was immensely influential on Christian theology. Averroes is shown on the ground in defeat as Aquinas is supposed to have claimed Aristotelianism into Christian theology from Islamic interpretations. (Lippo Memmi

(1291–1356)/Wikimedia Commons) Public Domain.

This does not mean that Europe evolved into a new society. Religion and superstition were still the norm. Heretics were regularly tortured and burned alive for going against church teachings. Giordano Bruno was famously burned alive by the Catholic Church for speculating on the existence of exoplanets (planets orbiting other stars). Galileo Galilei was placed under house arrest for the rest of his life for claiming that the Earth orbited the sun. However, seeds of knowledge had been planted, and a renaissance in thought was starting to emerge.

Media Attributions

- **Figure 1.9** Lippo Memmi – Triumph of St Thomas Aquinas – WGA15020 by Lippo Memmi (1291–1356) [Photo from the Web Gallery of Art], via Wikimedia Commons, is in the public domain.

A Renaissance in Thinking

Calvin says that he is certain, and [other sects] say that they are; Calvin says that they are wrong and wishes to judge them, and so do they. Who shall be judge? Who made Calvin the arbiter of all the sects, that he alone should kill? He has the Word of God and so have they. If the matter is certain, to whom is it so? To Calvin? But then why does he write so many books about manifest truth? . . . In view of the uncertainty we must define the heretic simply as one with whom we disagree. And if then we are going to kill heretics, the logical outcome will be a war of extermination, since each is sure of himself.

— Sebastian Castellio, *Concerning Heretics: Whether They Are to Be Persecuted* [quoted in Grayling (2007, pp. 53– 54)]

Where the Renaissance was the reawakening of the West to its ancient cultural and intellectual roots, the Scientific Revolution began as a critical response to ancient thinking and, in large part, that of Aristotle. This critical response was no quick refutation. Aristotle's physics might now strike us as quite naïve and simplistic, but that is only because every contemporary middle school student gets a thorough indoctrination in Newton's relatively recent way understanding of the physical world. The critical reaction to Aristotle that ignites the Scientific Revolution grew out of the tradition of painstakingly close study of Aristotle. The scholastic interpreters of Aristotle were not just wrongheaded folks stuck on the ideas of the past. They were setting the stage for new discoveries that could not have happened without their work. Again, our best critics are the ones who understand us the best and the one's from whom we stand to learn the most. In the Scientific Revolution, we see a beautiful example of Socratic dialectic operating at the level of the traditions of scholarship.

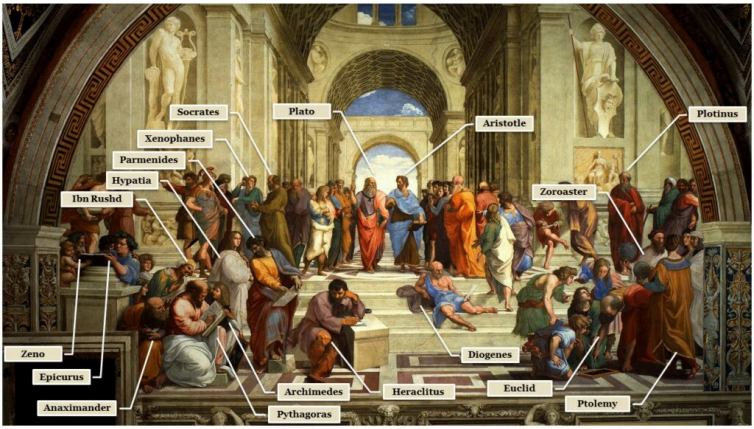


Figure 1.10 *The School of Athens by Raphael. This fresco in the Vatican shows the numerous philosophers who influenced Renaissance thinking showing the universality of philosophy.*

Europe also experienced significant internal changes in the 16th century that paved the way for its intellectual reawakening. In response to assorted challenges to the authority of the Catholic Church and the decadence of 16th century Catholic churchmen, Martin Luther launched the Reformation. The primary tenet of the reformation was that faith concerns the individual's relation to God who is knowable directly through the Bible without the intermediary of the Catholic Church. The Reformation and the many splintering branches of Protestant Christianity that it spawned undermined the dogmatic adherence to a specific belief system and opened the way for more free and open inquiry.

The undermining of Catholic orthodoxy brought on by the reformation combined with the rediscovery of ancient culture in the Renaissance jointly gave rise to the Scientific Revolution and what we often refer to as the Modern Classical period in philosophy. The reawakening of science and philosophy are arguably one and the same revolution. Developments in philosophy and science during this period are mutually informed, mutually influencing, and

intermingled. Individuals including Newton, Leibniz, and Descartes are significant contributors to both science and philosophy.

Media Attributions

- **Figure 1.10** Modified [cropped] image from "The School of Athens" [labeled] by Raphael (1483–1520), on Wikimedia Commons, is in the public domain.

The Scientific Revolution

Descartes first wrote the words, "*I think, therefore I am*", in his *Discourse on the Method* in 1637. What Descartes was trying to do with this simple phrase was actually much more ambitious than proving his own existence. He was trying to create a foundation for knowledge—a starting point upon which all future attempts at understanding and describing the world would depend. And, of course, to know anything requires a *knower* in the first place, right? In other words, we need a *mind*. Descartes recognized this, so he sought to found all knowledge on the existence of a thinker's own mind.

To prove that his mind existed, Descartes did not actually start with the act of thinking. He started instead with the act of radical doubt: he doubted his knowledge of the external world, God, and even his own mind. The reason for this radical doubt was to ensure that he did not accept anything without justification. So, the first stop was to doubt every single belief he held.

For one thing, Descartes understood that sometimes our senses deceive us, so we must doubt everything we experience and observe. He further understood that even experts sometimes make simple mistakes, so we should doubt everything we have been taught. Lastly, Descartes suggested that even our thoughts might be complete fabrications—be they the products of a dream or the machinations of an evil demon—so we should also doubt that those thoughts themselves are correct.

Yet in the process of doubting everything, he realized that there was something that he could not doubt—the very act of doubting! Thus, Descartes could at least be certain that he was doubting. But doubting is itself a form of thinking. Therefore, Descartes could also be certain that he was thinking. But surely there can be no thinking without a thinker (i.e., a thinking mind). Hence, Descartes proved to himself the existence of his own mind: I doubt; therefore, I think; therefore, I am.

Importantly for Descartes, this argument is intuitively true, as anyone would agree that doubting is a way of thinking and that thinking requires a mind. Thus, according to Descartes, the existence of a doubter's own mind is, for the doubter, established beyond any reasonable doubt.

It is important to notice what Descartes is claiming to exist through this line of reasoning. He is not suggesting that his physical body exists. So far, he has only proven to himself that his mind exists. If you think about it, the second "I" in "I think therefore I am" should really be "my mind"—hence, "I think, therefore *my mind* is". Descartes and others started a cascade of thinking that promoted originality of thought and openness to new ideas. Scholars throughout Europe engaged with each other across the next few decades, improving upon the ideas they inherited from past civilizations. This led to the Age of Enlightenment and the Scientific Revolution with Isaac Newton.

In 1687, Isaac Newton first published one of the most studied texts in the history and philosophy of science, *Philosophiæ Naturalis Principia Mathematica* (or the *Principia* for short). It is in this text that Newton first described the physical laws that are part and parcel of every first-year physics course, including his three laws of motion, his law of universal gravitation, and the laws of planetary motion. Of course, it would take several decades of debate and discussion for the community of the time to fully accept Newtonian physics. Nevertheless, by the 1760s, Newtonian cosmology and physics were accepted across Europe.

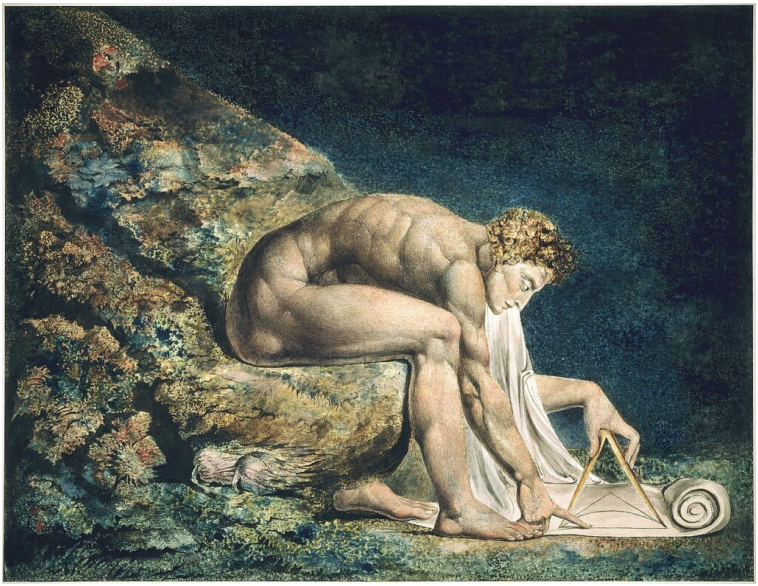


Figure 1.11 *Newton by William Blake. According to Blake, Newton made nature into a machine and thereby made it devoid of spirituality. (William Blake* (1757–1827)/Wikimedia Commons) Public Domain

Keats lightheartedly said Newton "has destroyed all the poetry of the rainbow, by reducing it to the prismatic colours" (Haydon, 1929). Indeed, Romanticism became a movement that challenged the Enlightenment and tried to emphasize emotion over intellectualism while glorifying the medieval past. William Blake famously painted Newton as engaged in petty mathematics and ignoring the true beauty of nature. However, as Dawkins (1998) later stated, Newton did not destroy the beauty of nature. Rather, he allowed us to see it with renewed eyes. As we will see in the rest of this book, science allows us to appreciate nature on many different levels and from a more nuanced perspective.

Media Attributions

- **Figure 1.11** Newton-WilliamBlake, by William Blake (1757–1827) [Photo from The William Blake Archive], via Wikimedia Commons, is in the public domain.

Summary

In this chapter, we explored the history of philosophy and how it developed across time in various cultures. A major part of the development of philosophy has been the application of critical thinking to subjects that had previously been the purview of religion and mythology. We saw that there were specific sociocultural and economic reasons for why philosophy as a tradition emerged in three cultures who then merged and exchanged ideas that led to modern philosophy and science. This shows us that it may be a mistake to divide philosophy and science into Eastern and Western ways of thinking or to relegate them to particular cultures. What we have today is the result of a long process of exchanges and dialogues between thinkers from all cultures. We are also part of this *Great Conversation* as we continue to wrestle with the hard questions of life, meaning, and the universe.

Key Takeaways

- Philosophy is a mode of thought that emerged in three different civilizations. It was different from what came before it in the form of mythopoeic thought.
- Ancient Greek, Indian, and Chinese civilizations have given birth to numerous schools of thought that challenged existing traditions. These, in turn, influenced other traditions in the Middle East and North Africa as well as Western Europe.
- Multicultural interactions and trade were

fundamental to the emergence of new ways of thinking.

- The intellectual changes that occurred in Western Europe during the 15th century led to the Renaissance and the Age of Enlightenment. These, in turn, eventually developed into the Scientific Revolution and the emergence of modern science.

Acknowledgements

Parts of this chapter were adapted from the following Open Education Resources:

- Russ Payne, W. (2015). *An Introduction to Philosophy*. BC Campus. https://collection.bccampus.ca/textbooks/an-introduction-to-philosophy-24/
- Barseghyan, H., Overgaard, N., & Rupik, G. (2018) Introduction to History and Philosophy of Science. Open Library, eCampus Ontario. https://ecampusontario.pressbooks.pub/introhps/

References

Dazzi, C., & Pedrabissi, L. (2009). Graphology and personality: An empirical study on validity of handwriting analysis. *Psychological Reports*, 105(3 Pt2), 1255–1268.

Downey, J. E. (1919). *Graphology and the psychology of handwriting.* Baltimore, MD: Warwick & York.

LaBianca, J., & Gibson, B. (2020). Here's what your handwriting says about you. Reader's Digest. Retrieved from https://www.rd.com/advice/work-career/handwriting-analysis

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2010). *50 Great myths of popular psychology: Shattering widespread misconceptions about human behaviour.* Chichester, England: Wiley-Blackwell.

Münsterberg, H. (1915). *Business psychology*. Chicago, IL: LaSalle Extension University.

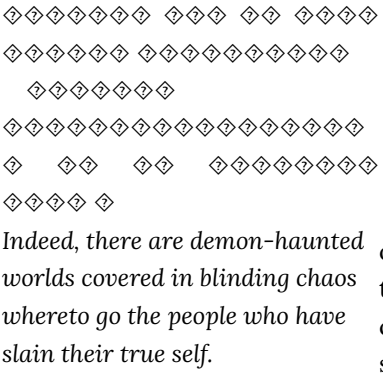
Wade, C. (1995). Using writing to develop and assess critical thinking. *Teaching of Psychology*, 22(1), 24–28.

II. DEMON-HAUNTED WORLDS

Learning Objectives

- Describe the concept of absolute knowledge and fallibilism
- Describe the problems with induction, sensation, and theory-ladeness
- Differentiate between the formal and empirical sciences

Introduction



Indeed, there are demon-haunted worlds covered in blinding chaos whereto go the people who have slain their true self.

— *Isha Upanishad* 3

The above quote was the inspiration for the astrophysicist Carl Sagan's 1995 book *The Demon-Haunted World: Science as a Candle in the Dark*. The book is an overview of the scientific method that encourages people to learn critical and skeptical thinking skills.

"Science has established beyond any reasonable doubt…", "physicists have proven that…", "it is absolutely true that…"—such phrases are abundant not only in popular science but also in academic literature. But is there such a thing as *proof* in science? Can we ever establish anything *beyond any reasonable doubt*?

The history of science shows us that scientific theories change through time. What was accepted 500, 250, or even 10 years ago may or may not still be accepted nowadays. In less than half a millennium, we moved from the theory of a finite geocentric universe to our contemporary cosmology. The theory of evolution by natural selection has been accepted for no more than a century. Our phylogenetic tree of life is constantly changing as we discover new species or new relations between species.

Theories in all fields of inquiry change over time. But if our theories change through time, then is there anything unchangeable in the mosaic? In other words: *Can we know anything with* absolute *certainty*? Or, alternatively: *Is there* absolute *knowledge*? This is the central question of this chapter.

Absolute Knowledge

It is sometimes obvious to us that something is evidently true. However, how do we know that something is true with absolute certainty. Let us consider some examples.

Example 1: 1 + 2 = 3

Here is the first one:

$$1 + 2 = 3$$

Question: How can we show that this proposition is actually true? Keep in mind, what we are asking here is not how we as human beings historically came to discover that one plus two equals three. Instead, the question is how do we *know* this proposition is true? How is it *justified*?

While it might be argued that it probably took centuries of human experience with the world before this proposition occurred to our ancestors, it could be argued that we do not need any experiments or observations to *justify* this proposition. Indeed, we do not need to conduct any experiments or observations in order to establish that one plus two actually equals three at all times. But if it is not based on experience, then what is it based on? The answer is: the truth of this proposition stems from the very meanings of the terms "two," "three," "plus," and "equals." What do we mean by "two"? Setting aside the question of a precise mathematical definition of the concept, we can say that "two" roughly means something like "one and another one." The same goes for the concept of "three," which is essentially short for "one, another one, and another one." But if that is what we mean by "two" and "three," then it follows that one plus two equals three.

This kind of reasoning is called deduction. In other words, this proposition holds true as long as "two," "three," and other terms

are understood in the way that they are usually understood. Importantly, we do not need to conduct any experiments or observations to ascertain that one plus two equals three at all times and in all places, for the proposition is true by virtue of the definitions of its terms.

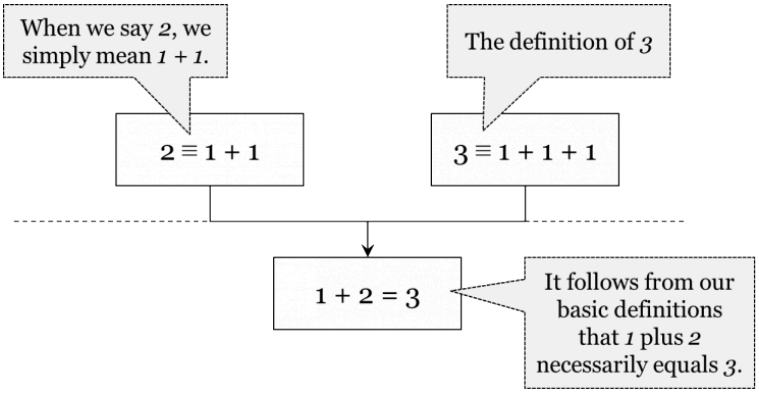


Figure 2.1 *Number definitions (Barseghyan et al./eCampusOntario)* CC BY 4.0

Example 2: All Swans Are White

Now, consider another example:

All swans are white.

How do we know this? Once again, the question is not when and under what circumstances humans first came to appreciate the truth of the statement. Rather, the question is how can this proposition be justified? What makes it true?

The intuitive answer is that this should have to do with experience; surely, one cannot know anything about the colour of swans unless one goes out and observes some actual swans! To establish that all swans are white, one should first observe the colours of individual swans and record the results of these observations. Now, suppose we go out and find a white swan. We

formulate this in a singular proposition that describes our experience:

Swan a *is white.*

We then find two more swans, observe that they are also white, and record this in the following two propositions:

Swan b *is white.*

Swan c *is white.*

Based on the results of these observations, we generalize and conclude that all swans are white. This inference from individual instances to a general proposition is called *induction*.

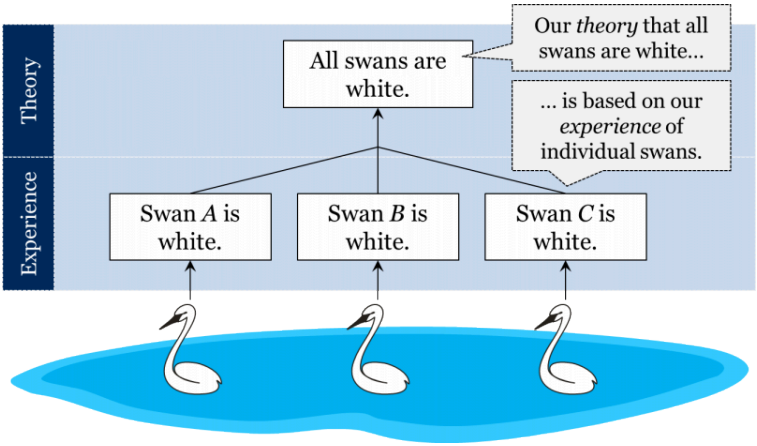


Figure 2.2 *Swan theory and experience (Barseghyan et al./eCampusOntario)* CC BY 4.0

In short, we seem to agree that the proposition "all swans are white" is justified by *experience*: it is an inductive generalization of the results of our *observations* of individual swans.

Example 3: The Law of Universal Gravitation

Our third and final example is from physics—the law of universal gravitation:

$$\begin{gathered} F=G\frac{m_1m_2}{r^2} \end{gathered}$$

Here, F stands for the force of attraction between two material objects, $m1$ and $m2$ are the respective masses of these two objects, r is the distance between them, and G is a constant that we can ignore here. The law says that there is a force of attraction between any two objects in the universe: the greater the masses of these objects, the greater the force of attraction between them; the greater the distance between them, the smaller the force. In other words, the attraction between two objects increases with mass and decreases with distance. The same idea can be expressed in more technical terms: the force of gravity is proportional to the product of the two masses and is inversely proportional to the square of the distance.

Now, how can we justify this proposition? How do we know it is true? More specifically, can the law of gravity be justified independently of experience, simply by virtue of the definitions of its terms, or does it take some experiments and observations to justify it? Is it more similar to "one plus two equals three" or "all swans are white"? In other words, is it possible to show that this proposition is true simply by analysing the definitions of "force," "mass," and "distance," or should we actually go out there and see how objects in the world behave in order to justify the law?

If we pay attention to the concepts of "force," "mass," and "distance," there is nothing there that suggests that all objects should attract each other with a force proportional to the product of the masses and inversely proportional to the square of the distance. In fact, we could conceive of an infinite number of different propositions which use the exact same concepts of "force," "mass," and "distance." For instance, we could say that the force of gravity is proportional to the *sum* of the masses, or that it is

inversely proportional to the *cube* of the distance, or that it *increases* with distance:

\begin{gathered} F=G\frac{m_1m_2}{r^2} \end{gathered}

Thus, there is no way to deduce the law of gravity merely from the definitions of its concepts, for there is clearly more than one way in which these concepts could be put together. But then how do we know which one of these numerous possibilities is true? The intuitive answer is that we have to go out and observe: there is simply no other way to know which one of these is the case without conducting experiments and observations. How might we do it? Perhaps if we were to observe that it holds for the Earth and a falling apple and also to observe that it holds for the Earth and the Moon, the Sun and Jupiter, and any two planets, then we could generalize and conclude that it holds between any two objects in the universe.

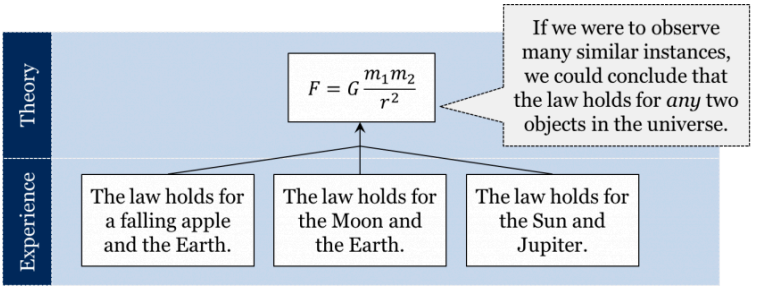


Figure 2.3 *Law of universal gravitational theory and experience (Barseghyan et al./eCampusOntario)* CC BY 4.0

What Does All This Mean?

In brief, the mode of justification of the law of gravity appears to be intuitively similar to that of "all swans are white." The fact that the law of gravity can be *expressed* as a mathematical equation should not confuse us here. It makes no difference how the proposition is expressed: the same proposition can be expressed mathematically

or in a natural language, but that does not change the gist of the proposition. The same proposition can be expressed as "1 + 2 = 3" or as "one plus two equals three"—its truth follows from the definitions of its concepts regardless of how we express it. Similarly, we have seen how the law of gravity could be expressed both as a mathematical equation and as "the force of gravity is proportional to..."—it makes no difference. The important point here is that one needs to conduct experiments and observations to check if the law is true.

Media Attributions

Unless otherwise noted, all figures are from the chapter "Absolute Knowledge" in the book *Introduction to History and Philosophy of Science* by Barseghyan et al., via eCampusOntario and are used under a CC BY 4.0 license.

Analytic vs. Synthetic

Let us now appreciate that the three propositions we have discussed belong to two different categories. We say that the proposition "one plus two equals three" is an *analytic* proposition, while the law of gravity and "all swans are white" are *synthetic* propositions. Analytic and synthetic propositions are two distinct types of propositions.

Now, what makes a proposition analytic? Analytic propositions are either definitions or deducible from definitions. Since analytic propositions are true by definition, they can never contradict the results of experiments and observations. For instance, the proposition "1 + 2 = 3" will hold at all times and places regardless of what we happen to observe.

Consider another typical analytic proposition: *All bachelors are unmarried.* This proposition is true by definition, since "bachelor" is *defined* as someone unmarried. If someone were to claim to have observed a married bachelor, we would probably conclude that she uses the word "bachelor" differently, for it is impossible to observe a married bachelor, insofar as we stick to our common definition of "bachelor." To express the same idea differently: an analytic proposition necessarily holds in all possible worlds (i.e., its opposite is inconceivable).

Thus, the proposition expressed as "1 + 2 = 3" is true in all possible worlds, in the sense that it is impossible to conceive of a world where this is not the case. Of course, we can always decide to change the definitions of our concepts and, for example, agree that "2" stands for "one, another one, and another one." But that would effectively change the meaning of "1 + 2 = 3"; it would no longer be the same proposition. So as long as "2," "3," "+," and "=" are defined the way they are commonly defined, the proposition "1 + 2 = 3" holds, come what may.

This is clearly not the case for synthetic propositions. Synthetic propositions are ones that are *not* deducible merely from the

definitions of their concepts. Because they are not deducible from the definitions of their concepts alone, they can potentially contradict the results of observations and experiments. This is the same as saying that synthetic propositions do not hold in all possible worlds, for their opposites are conceivable/imaginable. Thus, we can easily imagine a world where at least some swans are not white or where the law of gravity is different.

Table 2.1 Analytic vs Synthetic Propositions

Analytic Propositions	**Synthetic Propositions**
Deducible from definitions.	Not deducible from definitions.
Cannot contradict the results of experiments or observations.	Can contradict the results of experiments or observations.
Necessarily hold in all possible worlds: the opposite is inconceivable.	Do not necessarily hold in all possible worlds: the opposite is conceivable.

Here are a few additional examples of analytic propositions:

- $a\,(b + c) = ab + ac$
- *A centaur is a creature that is part human and part horse.*
- *The sum of the squares of the lengths of the sides of a triangle is equal to the square of the length of its hypotenuse:* $a2 + b2 = c2$.

All of the above propositions are such that their opposites are simply inconceivable, since they are either definitions or logically follow from definitions.

Here are some examples of synthetic propositions:

- *In combustion, methane combines with oxygen to form carbon dioxide and water:* $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$.
- *The game of football is played in four 15-minute-long quarters.*
- *Paris is the capital of France.*
- *There are no centaurs on planet Earth.*

All these propositions are attempting to say something about the

world we happen to inhabit and, consequently, they may or may not hold in other possible worlds. We can easily conceive of worlds where these propositions are false. Thus, we can imagine a strange world full of centaurs or a bizarre world where Paris is the capital of England. We can also conceive of an absolutely ridiculous world where the game of American football is played in two 45-minute halves.

How to Tell the Difference?

This brings us to a useful rule of thumb for distinguishing analytic from synthetic propositions:

- If the opposite of a proposition is conceivable (i.e., if it does not lead to logical contradictions) then the proposition is synthetic.
- If the opposite of a proposition is inconceivable (i.e., if it contains logical contradictions such as the idea of a married bachelor), then the proposition is analytic.

Knowing the Difference

Consider the following statements and see if you can decide which is analytic and which is synthetic:

- All bachelors are unmarried.
- All bachelors are lonely.
- All living beings have a heart and kidney.

- All squares have four sides.
- Forty-eight is a square number.

Formal vs. Empirical Sciences

The distinction between analytic and synthetic propositions comes in handy when distinguishing between *formal* and *empirical* sciences. While in *empirical* sciences—such as physics, chemistry, biology, psychology, sociology, and economics—there are both analytic and synthetic propositions, in *formal* sciences—such as logic or mathematics—*all* propositions are analytic. It is the absence of synthetic propositions that characterizes *formal* sciences; indeed, all propositions of math or logic are either definitions or follow from definitions.

Even those mathematical propositions that do not really strike us as analytic are essentially analytic. Consider, for instance, Euclid's famous postulate:

The sum of the angles of a triangle is two right angles.

It is not quite obvious that this postulate is analytic, since it is possible to negate it. In fact, Euclid's postulate has no place in the non-Euclidean *hyperbolic* geometry of Lobachevsky, where the sum of the angles of a triangle is less than 180°, as well as in the non-Euclidean *elliptical* geometry of Riemann, where the sum of the angles of a triangle is greater than 180°:

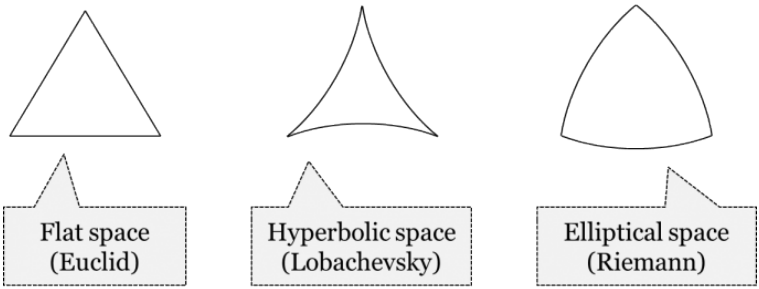


Figure 2.4 *Triangles in different geometries (Barseghyan et al./eCampusOntario)* CC BY 4.0

But if it is possible to negate some mathematical propositions, then in what sense are they said to be *analytic* propositions? The answer is this: while some mathematical propositions do not usually strike us as definitions, they essentially *define* properties of mathematical objects that are studied by that specific mathematical theory and, thus, are veiled definitions. What makes this possible is the fact that the objects of mathematical theories (such as triangles) are all *formal* (i.e., they do not exist *independently* of those mathematical theories that define them with their postulates and definitions). Thus, Euclid's postulate remains an analytic proposition in the context of Euclidean geometry, whereas some other postulates would be analytic in Lobachevsky's and Riemann's non-Euclidean geometries. The same goes for the objects of any formal science.

In contrast, theories of empirical science do not have the luxury of *defining* the properties of their objects. The objects of empirical science—natural or social—are all generally assumed to exist independently of the empirical theories that attempt to describe them. Thus, physical processes described by a physical theory are not defined/created by the physical theory. Similarly, a biological theory attempts to *describe* plants and animals that exist independently of that theory—the biological theory does not *define/create* organisms. By the same token, theories in political science try to *describe* political processes, like the succession of a new monarch, which would have occurred even if there was no political science.

Because the objects of empirical theories are not defined by empirical theories themselves, empirical theories cannot consist only of analytic propositions. They must contain at least some synthetic propositions that are hypotheses about the objects under study. For example, a physical theory cannot merely give us the definitions of "force of gravity," "mass," or "distance," without telling us how mass and distance affect the force of gravity. Similarly, a biological theory does not merely tell us what it means by "species" or "evolution"; instead, it tells us how species evolve through time. It is these relations between different features of natural or social

objects (structures, systems, processes, etc.) that any empirical theory attempts to uncover. Thus, hypotheses about these objects (i.e., synthetic propositions) are inevitable in empirical science.

How to Tell the Difference?

The rule of thumb for distinguishing formal from empirical theories is this:

- If a theory contains no synthetic propositions, it is a formal theory.
- If a theory contains at least one synthetic proposition, it is an empirical theory.

If we consider any theory from any one of the plethora of empirical sciences, we will see that it contains some synthetic propositions (i.e., propositions the opposite of which is conceivable). We have already seen this in the example of the law of gravity. The same applies to the chemical formula for the combustion of methane—it is possible to conceive of a world with a different set of chemical laws that make methane combust differently, or even not combust at all. The same applies to all synthetic propositions of all empirical sciences.

Media Attributions

Two Questions on Absolute Knowledge

Now that we understand the difference between analytic and synthetic propositions, we must specify the main question of this chapter for the two types of propositions:

- Can analytic propositions be *absolutely certain*?
- Can synthetic propositions be *absolutely certain*?

Let us first consider the former. Is it possible to establish the truth of an analytic proposition beyond any doubt? For instance, can the propositions of math or logic be proven with absolute certainty? The short answer is "yes" because analytic propositions are either definitions or follow from definitions; thus, they unfold what is already, in a sense, implicit in the definitions. Insofar as we stick to a given set of definitions, any analytic proposition that follows from them will be absolutely true. For example, as long as "bachelor" is defined as "an unmarried man," the proposition "all bachelors are unmarried" will remain absolutely certain. Similarly, as long as we stick to the definitions (including the postulates) of Euclid's geometry, its propositions will always hold.

To Euclid or Not to Euclid

There is a common misconception in the literature concerning the status of Euclidean geometry nowadays. It is often said that Euclid's geometry was rejected and replaced by a non-Euclidean geometry when general relativity became accepted ca. 1920. The error here is in the

confusion between *pure geometry* and *physical geometry*. As a *purely formal* theory, any geometry defines its own space as a *formal* object and then provides a description of that space; it does not attempt to say anything about the actual space of our universe.

Naturally, the spaces of different geometric theories can be very different: some geometries may define a flat space (i.e., a space where the sum of the angles of any possible triangle is 180°), while others may define a curved space (i.e., a space where this relationship for triangles does not hold true). The resulting geometries will be very different but, importantly, each of them will only contain analytic propositions and will provide an absolutely true description *of its own formally defined space*. Thus, Euclid's geometry is still and will always be an absolutely certain description of space *as it is defined in Euclid's geometry*. As a formal theory, it does not attempt to accomplish anything more.

In contrast, *physical* geometry attempts to describe the properties of the *physical* space of our universe. For instance, it attempts to find out whether the space of our universe is flat or curved. Any theory of physics has some physical geometry built into it: for instance, Euclid's geometry was implicit in Newtonian physics, and Riemann's non-Euclidean geometry is built into general relativity. But when any geometry becomes part of a physical theory, it ceases to be a formal theory and becomes physical geometry. It now makes hypotheses about the space of the universe and, thus, no longer consists merely of analytic propositions. Such a physical geometry is an empirical theory and is not to be confused with a pure geometry, which is a formal theory.

> Thus, when we say that Euclidean geometry was rejected, we have to keep in mind that it was only rejected as an empirical theory of the space of our universe. As a formal theory, it has never been and could never be under any threat. This is because it consists exclusively of analytic propositions which, by definition, can never be in conflict with any observational results.

This is equally true for any analytic proposition and, consequently, any theory that consists of only analytic propositions. Therefore, the answer to our first question is "yes": there can be absolutely certain analytic propositions. As a result, we can legitimately speak of mathematicians or logicians *proving* a certain theorem beyond any reasonable doubt.

Now, what about *synthetic* propositions: can they ever be absolutely certain? Can our hypotheses about the workings of the world ever be proven beyond any reasonable doubt? In other words, can we have absolutely certain theories in physics, chemistry, biology, psychology, sociology, economics, etc.? The short answer to this question is "no." To appreciate this, we have to consider how synthetic propositions are justified and what problems their mode of justification faces.

As we have seen with the example of white swans and the law of gravity, synthetic propositions must somehow be based on experience. So intuitively, we are inclined to think that the justification of a synthetic proposition requires some experience—some experiments and/or observations. This goes for both *singular* synthetic propositions, such as "swan *a* is white", and *general* synthetic propositions, such as "*all* swans are white." Indeed, how else can we justify any of these synthetic propositions, unless we go out there and observe?

Intuitively, we are inclined to accept the following general template. In order to justify a synthetic proposition, we find an

object, let us call it $p1$, and observe that it has some property, q. We then find a similar object, $p2$, and observe that it has the same property q. We repeat this with more objects of the same class—more p's—and conclude that all objects of type p have property q.

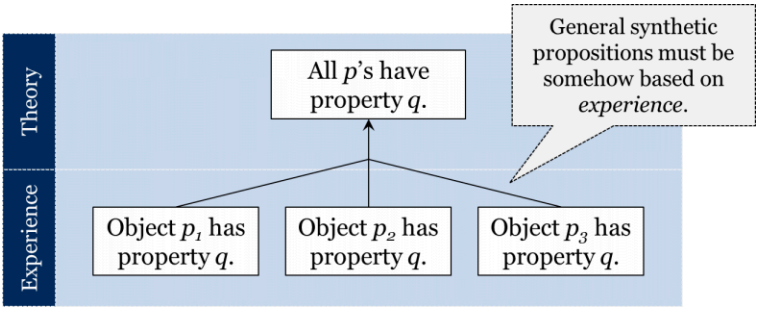


Figure 2.5 *Geneal synthetic propositions must be somehow based on experience (Barseghyan et al./eCampusOntario)* CC BY 4.0

But precisely because all synthetic propositions must eventually be based on experience, they cannot be absolutely certain. There are three giant obstacles that prevent synthetic propositions from being established beyond any reasonable doubt: *the problem of sensations*, *the problem of induction*, and *the problem of theory-ladenness*.

Problem 1: Sensations

First, let us distinguish between the human mind and the world outside of the human mind. Our feelings, emotions, sensations, theories, thoughts, beliefs, ideas, etc. are all in the mind. However, plants, animals, other people, and the Earth itself are in the external world, as they exist in reality and independently of the human mind. So, our sensations of swans will always be in our minds. Swans

as they exist "out there," objectively, should not be confused with my sensations of swans, which exist in our minds. In other words, things as they exist in reality should not be confused with things as we perceive them.

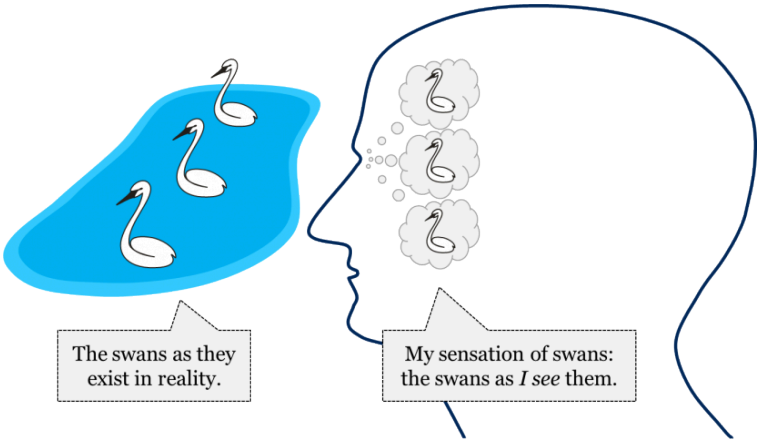


Figure 2.6 *Swans in reality vs. how we see them (Barseghyan et al./eCampusOntario)* CC BY 4.0

In order to claim that the proposition "all swans are white" is absolutely certain, we have to first establish that our sensations of swans are absolutely trustworthy: we have to make sure that swans as we see them are exactly the same as swans as they exist independently of the human mind. But how can we make sure that swans as we see them are exactly as they are in reality? How can we ever ascertain that our sensations convey the exact image of things as they really are?

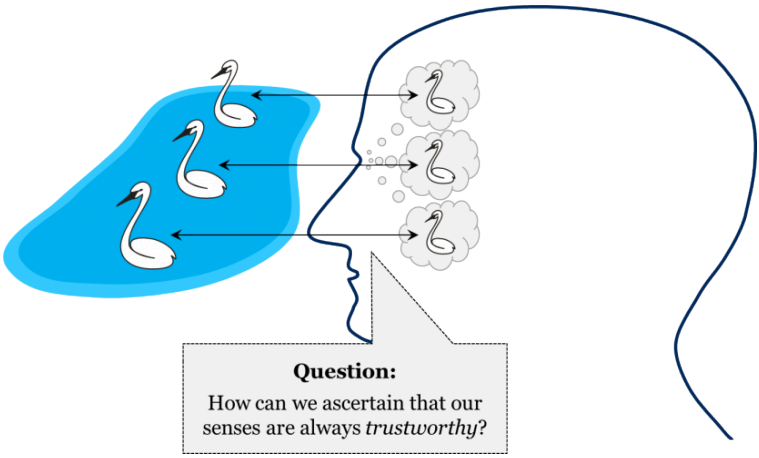


Figure 2.7 *How can we ascertain that our sensations of swans are trustworthy?* (*Barseghyan et al./eCampusOntario*) CC BY 4.0

There are reasons to suspect that our senses are not always trustworthy; surely, we do not normally trust everything we perceive when we are not sober! So, the question of the trustworthiness of our senses is far from idle; if we are going to claim that any of our hypotheses about the external world are absolutely certain, we have to prove that our sensations convey to our minds exact images of the things as they really are.

Suppose I see a cup of coffee in front of me. This is my visual sensation of a cup of coffee. How can I possibly make sure that there is *really* a cup of coffee in front of me? I could approach the cup, pick it up, smell it, and taste it. In other words, I could try to confirm my visual sensations with my sensations of touch, smell, and taste. Now, suppose that my sensations of touch, smell, and taste also suggest that there is a cup of coffee in front of me. Will this guarantee that there is *really* a cup of coffee in front of me? Of course not! All it would demonstrate is that my visual, tactile, olfactory, and gustatory sensations *cohere* with one another. That is all! But the question was not whether my sensations are

in agreement with one another. The question was whether my sensations are absolutely trustworthy.

In order to be absolutely sure that the cup of coffee is real, I would somehow have to grasp what the cup is like using a means that *bypassed* my senses. But that is impossible! We do not have a way of getting in touch with the world without using our senses. So even when *all* my senses seem to suggest that there is a cup of coffee in front me, all it really tells me is that *I sense* that there is a cup of coffee in front of me.

What if we asked other people to confirm the presence of the cup? Would that solve the problem? No, it would not. After they observed the cup, they would have to somehow convey the results of their observations to me. There is no way for me to receive their message that does not somehow involve my senses. On top of the fact that other people are relying on their senses, I have to rely on my senses to communicate with them about the information they gathered through their senses. So, all this would tell me is that I have an additional sensation that agrees with my other sensations.

The same goes for any tools we might want to use to confirm our sensations. Suppose I was to use some kind of coffee detector to ascertain that there is a cup of coffee in front of me. Say, the coffee detector had a screen that would display the message "this is coffee." Would that solve the problem? No, since I would have to read the screen or otherwise interact with the detector and the only way I can do that is though my senses—visual, auditory, etc. Therefore, I would only have one more sensation that is coherent with my other sensations.

In short, in order to be in a position to say that there is a cup of coffee in front of me, I have to be absolutely sure that my senses conveyed the exact picture of the cup as it really is. Yet this cannot be guaranteed. All I can check is that my sensations are in agreement with one another, but that does not necessarily guarantee that the world out there is exactly as my sensations suggest. This is known as *the problem of sensations*: there is no

guarantee that senses convey the exact picture of things as they *really* are.

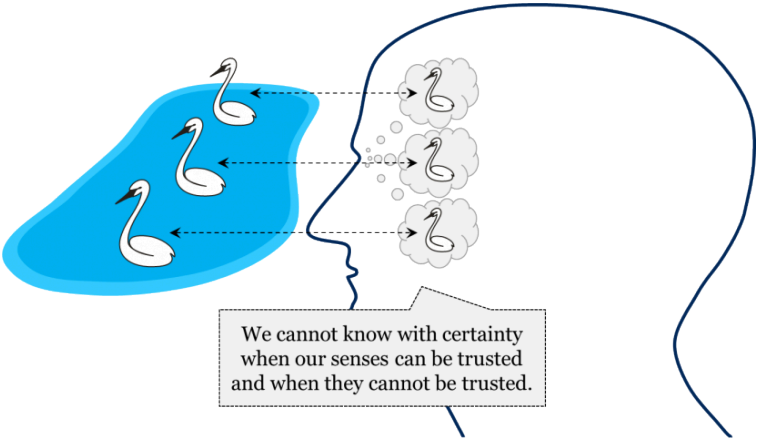


Figure 2.8 *There is no guarantee that we are seeing swans as they actually are (Barseghyan et al./eCampusOntario)* CC BY 4.0

It is even possible that we might be living in a computer simulation; we just do not have an absolutely certain way of knowing that we are not! While there have been many attempts to solve the problem of sensations, it is accepted nowadays that, strictly speaking, it cannot have a solution.

One suggested solution was to argue that because our senses are a product of evolutionary processes, they have adapted to perceive the world correctly, for otherwise humans would not survive. Is this a good solution? No, it is not, because it is based on the theory of natural selection, which is itself an empirical theory; the proposition "our senses adapt to their environment" is a synthetic proposition which can only be justified by invoking knowledge obtained with the use of our senses. Thus, the solution is itself circular: in order to show that our senses are trustworthy, it invokes the theory of natural selection, but in order to justify the theory of natural selection, we need to rely on evidence obtained with our senses.

Importantly, the problem of sensations in general does not mean that we cannot trust our senses; not trusting our senses would not be particularly conducive to our survival. The point here is *not* that our senses are *deceiving* us—no! The point is that we just cannot be *absolutely sure* that they are not deceiving us. There is a very important difference here: it is one thing to claim that senses are not trustworthy, but it is another thing to claim that we cannot know whether they are trustworthy or not. The problem of sensations is all about the latter: it tells us that we simply cannot be absolutely sure that our senses are not deceiving us.

Problem 2: Induction

Suppose for the sake of argument that there is no problem of sensations; suppose that our senses are absolutely trustworthy and convey the picture of the world as it really is. Even if we were to make this assumption, we would still be facing a serious problem. Anytime we observe something, the outcome is a *singular* synthetic proposition, such as "swan *a* is white" or "swan *b* is white."

Question: how can we ever prove any of our *general* synthetic propositions if the outcomes of our observations are always *singular* propositions? In other words, how is it possible to claim that *all* swans are white if our observations only tell us that *individual* swans we have observed so far are white? Is there any guarantee that there will not be any non-white swans out there (as black swans do exist in nature)?

Similarly, how can we ever be sure that the law of gravity has been proven beyond any doubt? As a general synthetic proposition, it must be based on observations, but all our observations concern individual cases. We can observe that the law holds between the Earth and the Sun, or the Sun and Jupiter, and so on, but does this prove that the law holds for *any* two objects in the universe? In

other words, how can we be sure that our *inductive* generalizations hold water?

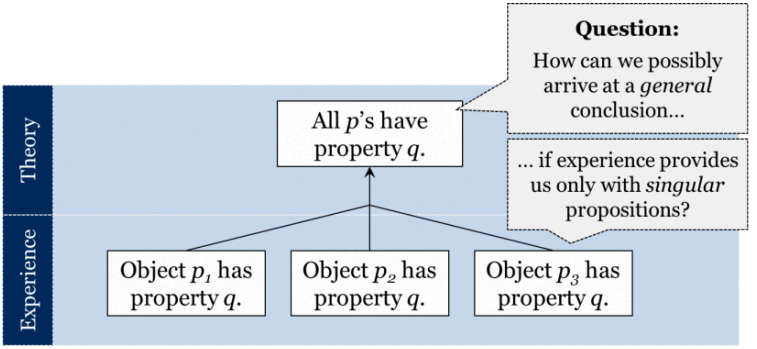


Figure 2.9 *How can we have a general conclusion if experience only has singular propositions? (Barseghyan et al./eCampusOntario)* CC BY 4.0

Because general propositions refer to all objects within a given class, they attempt to say something not only about those objects that we have already observed but also *any* object of that type. To establish that "all swans are white" beyond any reasonable doubt, we should make sure we have observed each and every swan out there, including all swans that have ever lived and all swans that will be born in the future. Surely, this is impossible! But if we are never in a position to observe all swans, how can we be absolutely sure that all swans are white? It would suffice to observe one counterexample—one non-white swan—and the whole generalization would fall apart. Thus, no matter how many millions of white swans we have managed to observe, there is no logical proof that the swans we will observe in the future will also be white.

Similarly, we can observe millions of instances of objects acting in accord with the law of gravity, but this does not prove that *all* objects in the universe obey this law. Even if all the physicists in the world were to collect their observational results in a huge database, the outcome would still be limited; it would never cover *all* objects in the universe.

This is the gist of *the problem of induction*: because our experience is always limited, our inductive generalizations are inevitably *fallible*.

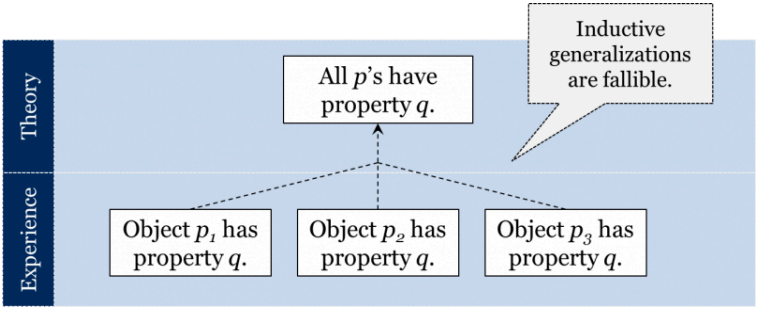


Figure 2.10 *Inductive generalizations are fallible (Barseghyan et al./eCampusOntario)* CC BY 4.0

The only exceptions here are those rare cases where the number of objects within a class is *finite* and we have managed to observe each and every one of them. For example, the proposition "all capitals of Rus' (a kingdom in medieval Europe) were located between 45° and 60° northern latitude" does not face the problem of induction since Rus' had a finite number of capitals and all of their locations are known. However, in most cases, there are so many objects within a class that we simply cannot observe all of them. In those numerous cases, our generalizations remain fallible because of the problem of induction.

Now, is there a way to solve the problem of induction? One classic attempt to solve the problem was to invoke the so-called *principle of the uniformity of nature*—the idea that objects of the same class behave similarly in similar circumstances. According to this principle, there are certain regularities in nature and identical causes always lead to identical effects. But since nature is uniform, so the argument goes, we do not need to observe millions of objects of the same class. It would suffice to observe a few objects of the same class, note what they have in common, and then safely

generalize that to *all* objects of that class. This generalization would hold because nature is uniform (i.e., because identical initial conditions always produce identical effects).

According to this line of reasoning, we do not need to observe all swans in the world to prove that all of them are white. They are all white because the ones we have observed so far have all been white and because the principle of uniformity tells us that all swans are similar as they are all products of the same biological mechanism. By the same token, there is no need to test the law of gravity for every pair of objects in the universe since, according to the principle of uniformity, all objects with mass must act similarly. In short, the principle allows us to safely extrapolate from past experience to all cases of the same class and to ensure that there will not be any surprises in the future.

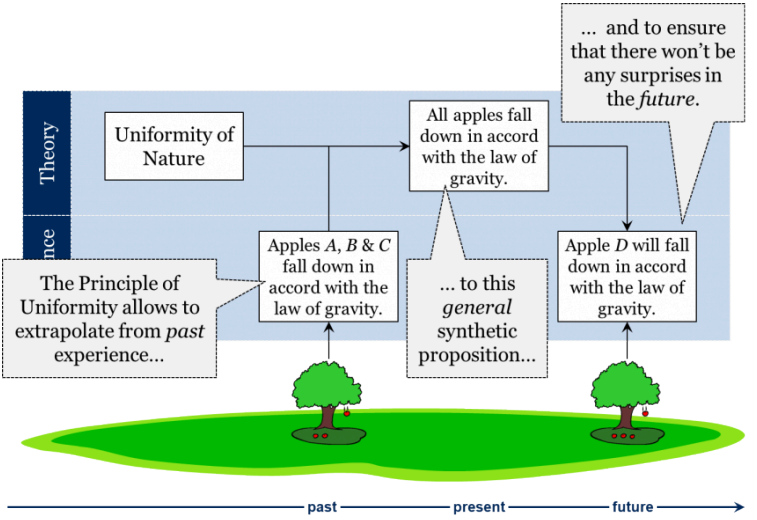


Figure 2.11 *The Principle of uniformity applies to the law of universal gravitation (Barseghyan et al./eCampusOntario)* CC BY 4.0

With this wonderful principle of the uniformity of nature, we are no

longer required to make an infinite number of observations to prove a general synthetic proposition.

Now, does this solution hold water? It is true that if the principle of uniformity were established beyond any reasonable doubt, then it would solve our problem. But how do we know that nature is regular and uniform? How is the principle of uniformity itself justified?

There are two options here:

> **Option 1**: We can try to argue that the principle itself is an *analytic* proposition (i.e., that it is true by definition). But this is not a viable option: the principle of uniformity is not an analytic proposition, since its opposite is conceivable. For instance, we can easily conceive of a world where regularities allow for exceptions. We can equally conceive of a world with no regularities whatsoever. Clearly, the principle of uniformity is itself a synthetic proposition—it is a hypothesis about the workings of *our* world.

This leaves the second option:

> **Option 2**: We can try to argue that, as a general synthetic proposition, the principle of uniformity is justified through thousands of years of human experience. We see regularities everywhere around us, and we see that these regularities do not change through time: apples have been falling to the ground since time immemorial and they still fall down just fine. Thus, the principle of uniformity is based on our observations that objects of the same type have always behaved similarly in similar conditions. In other words, the principle is itself an inductive generalization of our past experience: it assumes that objects of the same type will always behave similarly because that is how they have behaved so far. Thus, the principle of uniformity faces the same problem as any other general synthetic proposition—the problem of induction.

Therefore, we cannot solve the problem of induction by means of the principle of uniformity because that would lead to a vicious circle: we would use the principle of uniformity to guarantee the

absolute certainty of our inductive generalizations, but we would also use induction to justify the principle of uniformity itself, which we would then use to justify induction, which would then justify the principle of uniformity, and so on. We would end up in a vicious circle!

Nowadays, it is accepted that the problem of induction does not have a perfect solution, and therefore, our inductive generalizations based on limited experience are fallible. Of course, this does not mean we have to stop making inductive generalizations; all it suggests is that these generalizations are not absolutely certain. We cannot and should not avoid drawing general conclusions. We just have to appreciate that they are not proven beyond any reasonable doubt.

Problem 3: Theory-Ladenness

Let us now proceed to our third problem—*the problem of the theory-ladenness of observations*. Consider the following image:

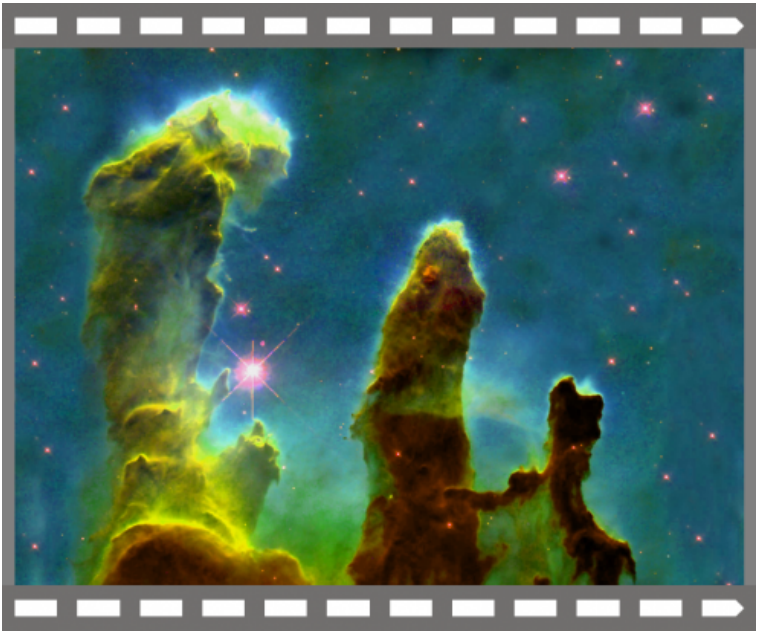


Figure 2.12 *What do you see? (Barseghyan et al./eCampusOntario)* CC BY 4.0

What do we see here? Those of us without astronomical training will probably see some stars obscured by what seems to be a cloud of poisonous gas. However, people trained in astronomy would most likely say that this is a photograph of the Pillars of Creation, which is a region of interstellar gas and dust about 7,000 light-years from Earth in the Eagle Nebula. In other words, our observation of the Pillars of Creation is *laden* with our astronomical theories. This phenomenon is known as *the theory-ladenness of observations*.

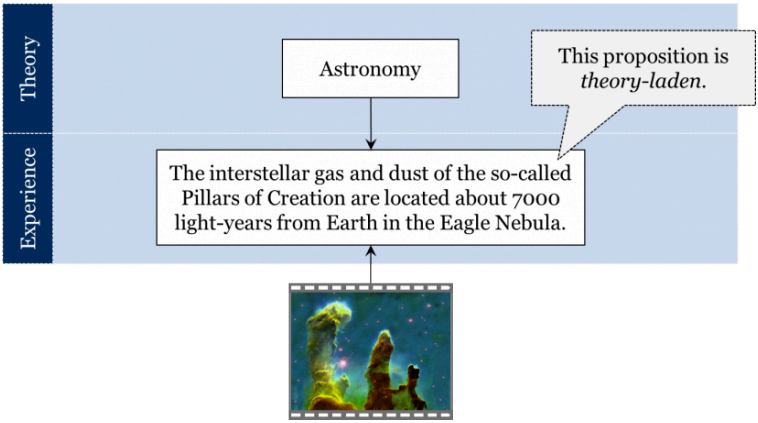


Figure 2.13 *The Pillars of Creation is theory-laden (Barseghyan et al./eCampusOntario)* CC BY 4.0

Nowadays, it is accepted that all observations are *theory-laden* (i.e., they depend on some accepted theories). Another way of saying the same thing is that there are no pure statements of fact, in the sense that all statements of fact presuppose one theory or another. Note that the word "pure" here means "unaffected by any theory." The point is that no observational statement is pure: all observational statements are shaped by one theory or another.

Consider another image:

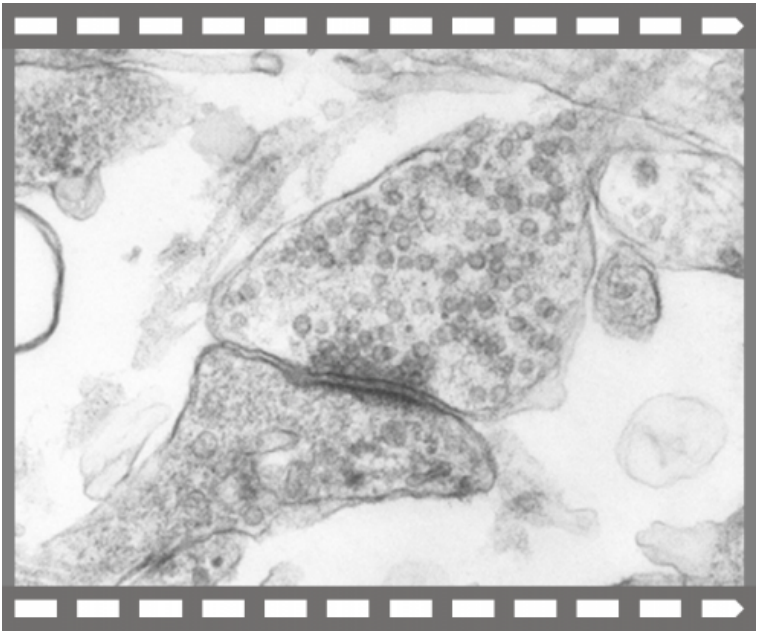


Figure 2.14 *What do you see? (Barseghyan et al./eCampusOntario)* CC BY 4.0

What is this? While a layman will probably see a fuzzy greyish picture, a person educated in neurophysiology will see a *synapse*. More specifically, an educated person will see a *presynaptic terminal* in the upper right portion of the image and a *postsynaptic terminal* in the lower left. They will recognize the presynaptic terminal by the small dark circles it contains. They will see these circles as packets of neurotransmitter molecules called *synaptic vesicles*. These vesicles can release neurotransmitters into the thin gap between the presynaptic terminal and the postsynaptic terminal, which is called the *synaptic cleft*. Neurotransmitter molecules can cross this gap and bind chemically to receptor molecules on the postsynaptic terminal. But this understanding is only possible if we accept some of the theories of neurophysiology.

This is another example of our accepted theories shaping the results of our observations.

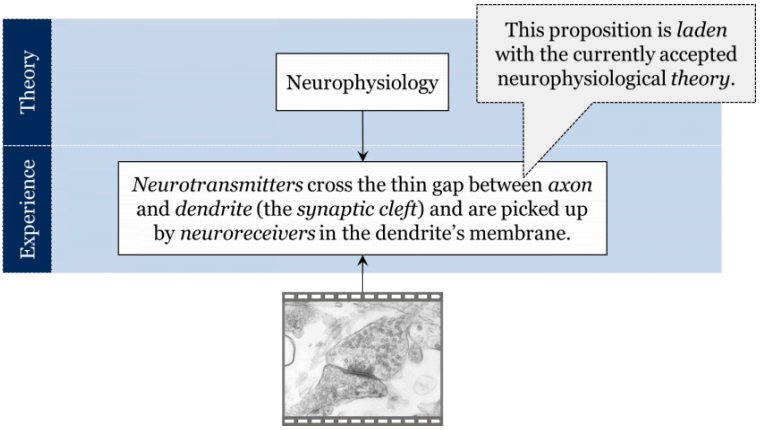


***Figure* 2.15** *Neurophysiology is also theory-laden (Barseghyan et al./eCampusOntario)* CC BY 4.0

The general point here is that all propositions that describe our experience are inevitably theory-laden: there are no pure statements of fact.

To appreciate this point, let us take another example. Suppose we look at the Moon through a telescope and observe that it has mountains. Why is this not a pure statement of fact? The proposition "there are mountains on the moon" is not a pure statement of fact because it assumes the reliability of the telescope. But the reason we rely on the telescope is because we know that it was constructed in accord with our accepted optical theory, which suggests that a certain combination of magnifiers produces a trustworthy image.

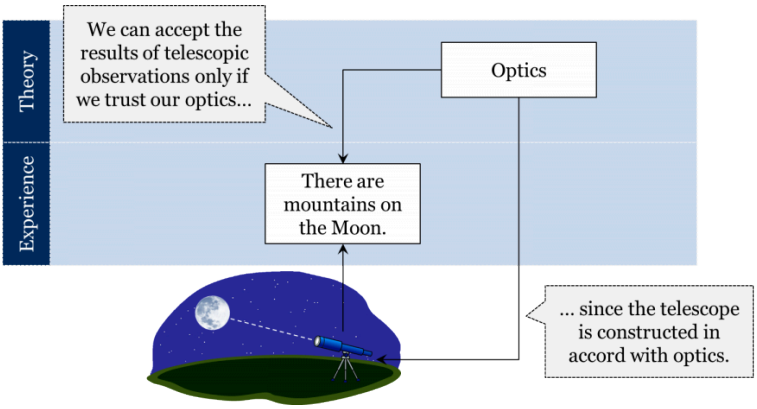


Figure 2.16 *To accept what we see through a telescope, we must trust the accepted optical theory that was used to build it (Barseghyan et al./eCampusOntario)* CC BY 4.0

Generally speaking, any observational instrument presupposes one theory or another, and therefore, the observational results obtained by means of an instrument depend on the theories in accordance with which the instrument is constructed.

But what about those cases where we do not seem to use any instruments? Suppose I look out the window with my naked eye and say, "It's raining." How is this proposition theory-laden? Surely, I am not using any instruments, so what theories are presupposed here? Even in this most simple case, my observation is based on some basic assumptions about my own physiology and about optics. In fact, we do not trust our observations at all times, but only when they satisfy certain basic criteria—such as proper illumination, the absence of visual obstacles (e.g., fog), not having taken a hallucinogenic drug, not being subject to hallucinations due to malnourishment or sickness, etc. For instance, when a teaspoon appears bent or broken in a glass of water, we do not trust this sensation for we have some prior knowledge of how these things work. We essentially rely on some basic experiential knowledge of optics.

Once again, it does not matter that these theories often remain unarticulated; what matters is that any observation—even a naked-eye observation—is based on some theoretical assumptions. At the very minimum, there are some physiological and optical assumptions that allow us to determine which sensations to trust and which not to trust. Thus, there are no pure statements of fact; all observations are theory-laden.

The problem of theory-ladenness becomes especially daunting once we appreciate that the theories through which we see the world change through time. Effectively, the same image can be interpreted differently depending on which set of theories we use to interpret it. Consider the example of a falling apple. Where Aristotelians would see a heavy object composed predominantly of the elements earth and water descending towards the centre of the universe, Newtonians would see a gravitational attraction between the Earth and the apple. In contrast, nowadays, we would interpret this as a case of inertial of motion in the curved space-time continuum produced by the Earth's mass. Similarly, where chemists of the mid-18th century would see dephlogisticated air, we would see oxygen. The same goes for any observation in any field of inquiry.

Media Attributions

Fallibilism

Let us recap. We have discussed three problems:

- **The problem of sensations:** There is no guarantee that our senses convey the exact picture of things as they really are.
- **The problem of induction:** No matter how many confirming instances are observed, inductive generalizations remain fallible.
- **The problem of theory-ladenness:** The results of experiments and observations are shaped by our accepted theories.

Together, these three problems suggest that synthetic propositions cannot be established beyond any reasonable doubt. Thus, we arrive at the conception of **fallibilism**, which holds that no synthetic proposition is infallible and, thus, empirical knowledge cannot be absolutely certain.

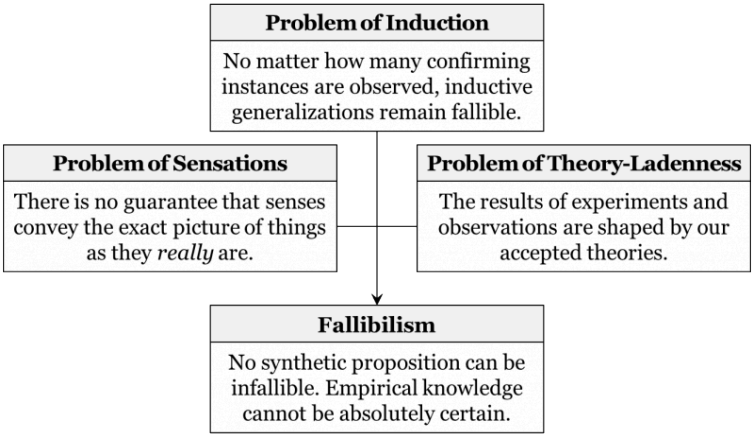


Figure 2.17 *The problems of sensations, induction, and theory-ladenness all connect to fallibilism (Barseghyan et al./eCampusOntario)* CC BY 4.0

Fallibilism is the currently accepted conception concerning the possibility of absolute empirical knowledge. However, just because science does not know everything with absolute certainty, it does not mean that it does not know anything (there are conclusions about nature that we do know to some degree of certainty). This is a common tactic used by those who oppose science and scientific evidence. This tactic of "Science doesn't know everything" is used to promote alternative ideas, such as New Age crystal therapy and other pseudoscientific nonsense. One technique to check whether something is scientific or not is to use Popper's (1959/2002) definition of falsifiability: falsifiability involves seeing whether a proposition could be rendered false under some testable circumstance. Sagan (1995) provides a wonderful example to illustrate this:

The Dragon in My Garage

"A fire-breathing dragon lives in my garage"

Suppose I seriously make such an assertion to you. Surely, you would want to check it out to see for yourself. There have been innumerable stories of dragons over the centuries, but no real evidence. What an opportunity!

"Show me," you say. I lead you to my garage. You look inside and see a ladder, empty paint cans, an old tricycle—but no dragon.

"Where's the dragon?" you ask.

"Oh, she's right here," I reply, waving vaguely. "I neglected to mention that she's an invisible dragon."

You propose spreading flour on the floor of the garage to capture the dragon's footprints.

"Good idea," I say, "but this dragon floats in the air."

Then you will use an infrared sensor to detect the invisible fire.

"Good idea, but the invisible fire is also heatless."

You will spray-paint the dragon and make her visible.

"Good idea, but she's an incorporeal dragon and the paint won't stick."

And so on. I counter every physical test you propose with a special explanation of why it will not work.

Now, what is the difference between an invisible, incorporeal, floating dragon who spits heatless fire and no dragon at all? If there is no way to disprove my contention, no conceivable experiment that would count against it, what does it mean to say that my dragon exists? Your inability to invalidate my hypothesis is not at all the same thing as proving it is true. Claims that cannot be tested and assertions immune to disproof are veridically worthless, whatever value they may have in inspiring us or in exciting our sense of wonder. What I am asking you to do comes down to believing my say-so in the absence of evidence.

Fallibilism opposes the conception of *infallibilism*, which holds that empirical science can provide absolute knowledge (i.e., that there can be absolutely certain synthetic propositions). Roughly speaking, infallibilism was accepted until about the early 20th century. Ever since the times of Aristotle and all the way into the early 20th century, it was generally assumed that at least some parts of

empirical science had been established once and for all. However, this has been shown to be wrong time and time again.

In the 19th century, August Comte formulated the idea of a hierarchy of science (Cole, 1983). His idea was that science develops over time from the most general scientific discipline (sciences that deal with observable nature, such as astronomy) to more complex systems. People started speculating about the stars and heavenly objects from the earliest of times and developed methods to predict eclipses and the weather. These developed into specifics about the nature of the physical worlds (physics) and the substances that make up the world (chemistry). In this view, biology is a more complicated chemistry and psychology is an even more complicated biology. The idea is that disciplines that are higher on this hierarchy are dependent on developments in their predecessors. We can see some evidence for that in how psychology developed in the late 19th century compared to physics, which developed much earlier. More recent developments in psychology have involved brain imaging techniques that are based on those developed within the fields of biology and physics.

This is one way of looking at the development of science. As we see in Figure 2.18, while astronomy is the basis of physics, it could also be placed at a higher level in the scientific hierarchy as it describes objects in the universe that are bigger in scale than biological or geological systems. In either sense, we should not mistake the idea of *complex science* as being "more difficult" or "better." *Complex*, in this context, means how many phenomena need to be taken into account to understand a system. Physicists can discuss subatomic particles and molecular bonds without reference to higher level entities such as animal cells. However, a biologist needs to consider chemical bonds and the physics behind them when studying biology, while a psychologist needs to take into account the biology of the brain and nervous system.

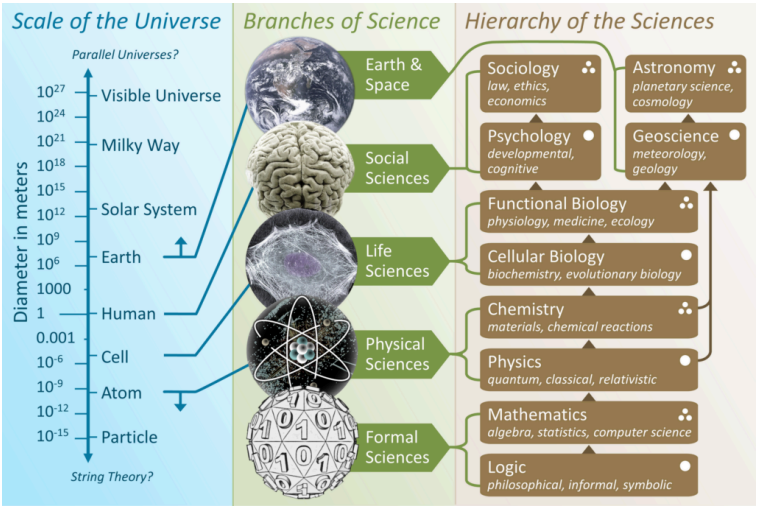


Figure 2.18 *The Hierarchy of Science mapped onto Scales of the Universe. We can see the formal sciences at the lower end while the empirical sciences are further up the hierarchy. We can see astronomy placed higher up because of its scale but we also see it is more closely related to physics than to the other sciences. (Efbrazil/Wikimedia Commons)*

Philosophers would debate as to how exactly there can be absolutely certain synthetic propositions and which propositions of science have or have not been established beyond any reasonable doubt. However, it was generally assumed that absolutely certain empirical knowledge is possible in one way or another. In the 18th and 19th centuries, the often-cited example of absolutely certain empirical theory was Newtonian physics. Around the time of its replacement with general relativity ca. 1920, it became clear that no empirical theory could ever be safe. This marked the transition to fallibilism. Table 2.2 will help summarize the debate between fallibilism and infallibilism.

Table 2.2 Can There Be Absolutely Certain Synthetic Propositions?

Yes	**No**
Infallibilism	Fallibilism
Synthetic propositions can be infallible. Empirical knowledge can be absolutely certain.	No synthetic proposition can be infallible. Empirical knowledge cannot be absolutely certain.
◈ Problem of sensation ◈ Problem of induction ◈ Problem of theory-ladenness	◈ Problem of sensation ◈ Problem of induction ◈ Problem of theory-ladenness

The table begins with a formulation of the *question* at issue (the topic): can there be absolutely certain synthetic propositions? It then presents the two opposing answers to the question—the two *conceptions* (i.e., positions, viewpoints). Finally, the table lists the *reasons* (pros and cons) for and against these conceptions. The table shows that there are good reasons to accept the conception of fallibilism, which is exactly what contemporary philosophy does.

It is important to repeat that fallibilism does not suggest that we have to stop relying on our senses, making inductive generalizations, or describing our observations by means of our accepted theories. In fact, all sciences still do this and that does not impede their advancement. The point here is that because we make observations and generalizations, our synthetic propositions are inevitably fallible (i.e., they are not absolutely true).

This brings us to our answers to the two questions we formulated at the outset of the chapter.

- Can analytic propositions be absolutely certain? *Yes.*
- Can synthetic propositions be absolutely certain? *No.*

This outcome has serious consequences for both empirical and formal sciences. Since only analytic propositions can be absolutely true, absolute knowledge is only achievable in formal sciences, such as mathematics or logic. In contrast, absolute certainty is unachievable in empirical sciences, since empirical theories always contain synthetic propositions which, as we know, cannot be

established beyond any reasonable doubt. Importantly, this is the reason why there cannot be such a thing as a proof in empirical science. Often scientists speak of having proven something in physics, biology, or even social science, but strictly speaking proof is only achievable in formal sciences; empirical sciences do not prove anything (Table 2.3).

Table 2.3 Formal vs. Empirical Science

Formal Science	Empirical Science
E.g. Logic, Mathematics, Systems Theory, game Theory, information Theory	**E.g.** Physics, Chemistry, Biology, Psychology, Sociology, Economics, Political Science
Contains only *analytic* propositions	Contains *synthetic* propositions
Absolute certainty is achievable	Absolute certainty is unachievable
Propositions can be demonstratively *proven*	Propositions can only be *confirmed* by experience

Now, if there is no such thing as a strict proof in empirical sciences, does this mean that every empirical theory is as good as any other? Are all competing empirical theories equally good, or is there a way to *compare* competing theories and find out which one of them is better? It is true that all empirical theories are affected by the problems of sensations, induction, and theory-ladenness; all empirical theories without exception are fallible. Yet we still find *some* of these empirical theories *acceptable*. Such ideas are fundamental to how scientists view the world. However, how do we educate ourselves and the public at large about which ideas or theories to accept? Do we all need to become philosophers of science to be able to engage with the world of science? Obviously, this is not the case.

In the next chapter, we will discuss science and non-science in greater detail. Before we get there, let us finish this chapter with a look at a toolkit that Sagan (1995) created to check whether something is empirically supported or not.

Carl Sagan's Baloney Detection Kit

Think of any idea in modern discourse that is being discussed. Use the following suggestions in thinking about that idea:

1. Wherever possible there must be independent confirmation of the "facts."
2. Encourage substantive debate on the evidence by knowledgeable proponents of all points of view.
3. Arguments from authority carry little weight.
4. Develop more than one hypothesis.
5. Try not to get overly attached to a hypothesis just because it's yours.
6. If there is a chain of argument, every link in the chain must work (including the premise)—not just most of them.
7. Choose the hypothesis which requires the least number of assumptions to work.
8. Always ask whether the hypothesis can be, at least in principle, falsified.

Media Attributions

- **Figure 2.17** is from the chapter "Absolute Knowledge" in the book *Introduction to History and Philosophy of Science* by Barseghyan et al., via eCampusOntario and is used under a CC BY 4.0 license.

- **Figure 2.18** "The Scientific Universe" by Efbrazil, via Wikimedia Commons, is used under a CC BY-SA 3.0 license.

Summary

In this chapter, we explored some fundamental ideas about absolute knowledge and whether this is possible to attain. We looked at problems with sensation, induction and theory-ladeness. We learned that fallibilism is the currently accepted conception concerning the possibility of absolute empirical knowledge and the differentiated between the formal and empirical sciences.

Key Takeaways

- One needs to conduct experiments and observations to check if a scientific law is true.
- Analytic propositions are either definitions or deducible from definitions. Since analytic propositions are true by definition, they can never contradict the results of experiments and observations.
- Synthetic propositions are ones that are not deducible merely from the definitions of their concepts. Because they are not deducible from the definitions of their concepts alone, they can potentially contradict the results of observations and experiments.
- In empirical sciences—such as physics, chemistry, biology, psychology, sociology, or economics—there are both analytic and synthetic propositions.
- In formal sciences, such as logic or mathematics, all propositions are analytic.

- Fallibilism means that absolute empirical knowledge is not possible.
- Fallibilism is the currently accepted conception concerning the possibility of absolute empirical knowledge. It opposes the conception of infallibilism, which holds that empirical science can provide absolute knowledge.

Acknowledgements

Parts of this chapter were adapted from the following Open Education Resources:

- Barseghyan, H., Overgaard, N., & Rupik, G. (2018). Introduction to History and Philosophy of Science. eCampusOntario. https://ecampusontario.pressbooks.pub/introhps/chapter/chapter-2-absolute-knowledge/

References

Cleland, C. E., Chute, R. D., & Haltiner, R. E. (1984). NAUB-COW-ZO-WIN discs from Northern Michigan. *Midcontinental Journal of Archaeology*, 9(2), 235–249. https://www.jstor.org/stable/20707933

Confucius. (1998). *The Analects* (D. C. Lau, Trans.). Penguin Books Ltd. (Original work published 447)

Dawkins, R. (1998). *Unweaving the rainbow: Science, delusion, and the appetite for wonder*. Houghton Mifflin.

Doniger, W. (2014). *On Hinduism*. Oxford University Press.

Flood, G. D. (1996). *An introduction to Hinduism*. Cambridge University Press.

Frankfort, H., Frankfort, H. A., Wilson, J. A., Jacobsen, T., & Irwin, W. A. (1946/1977). *The intellectual adventure of ancient Man: An essay on speculative thought in the ancient near east*. The University of Chicago Press.

Graeber, D., & Wengrow, D. (2023). *The dawn of everything: A new history of humanity*. Picador.

Grayling, A. C. (2007). *Toward the light of liberty: The struggles for freedom and rights that made the modern Western world*. Walker.

Haydon, B. (1929). Chapter XVII 1816–1817. In A. P. D. Penrose (Ed.), *The autobiography and memoirs of Benjamin Robert Haydon 1786–1846* (p. 231). Minton Balch & Company.

Lenik, E. J. (2012). The thunderbird motif in Northeastern Indian art. *Archaeology of Eastern North America*, 40, 163–185. https://www.jstor.org/stable/23265141

Liang-Hung, L. & Yu-Ling, H. (2009). Confucian dynamism, culture and ethical changes in Chinese societies – A comparative study of China, Taiwan, and Hong Kong. *The International Journal of Human Resource Management*, 20(11), 2402–2417. https://doi.org/10.1080/09585190903239757

Perrett, R. W. (Ed.). (2000). *Indian philosophy: A collection of readings: Vol. 3. Metaphysics*. Garland.

Wisdom, S. (2015). *Handbook of research on advancing critical thinking in higher education*. IGI Global.

III. PSEUDOSCIENCE

Learning Objectives

- Describe levels of scientific explanation
- Differentiate between science and non-science
- Define pseudoscience

Introduction

– Thirukkural 423

◇◇◇◇◇◇◇◇◇◇
◇◇◇◇◇◇◇◇ ◇◇◇◇◇◇
◇◇◇◇◇◇◇◇◇◇
◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇
◇◇◇◇◇◇ ◇◇◇◇◇◇ ◇◇◇◇◇◇
Regardless of where one heard about something;wisdom is being able to discern its true nature

Despite the differences in their interests, areas of study, and approaches, all scientists have one thing in common: they rely on scientific methods. Researchers use scientific methods to create new knowledge about the causes of behaviour, whereas practitioners, such as clinical-, counselling-, industrial/organizational-, and school-psychologists or field biologists and engineers use existing research to enhance the everyday life of others. Scientific research is important for both researchers and practitioners.

Of all of the sciences, psychology is probably the one that most non-scientists feel they know the most about. Because psychology is concerned with people and why they do what they do, we are all "intuitive" or "naive" psychologists. We rely on common sense, experience, and intuition in understanding why people do what they do. We all have an interest in asking and answering questions about our world, and in making sense of ourselves and other people. We want to know why things happen, when and if they are likely to happen again, and how to reproduce or change them. Such knowledge enables us to predict our own behaviour and that of others. We may even collect data (i.e., any information collected through formal observation or measurement) to aid us in this undertaking.

It has been argued that people are "everyday scientists" who conduct research projects to answer questions about behaviour (Nisbett & Ross, 1980). When we perform poorly on an important test, we try to understand what caused our failure to remember or

understand the material and what might help us do better the next time. When our good friends Monisha and Charlie break up, despite the fact that they appeared to have a relationship made in heaven, we try to determine what happened. When we contemplate the rise of terrorist acts around the world, we try to investigate the causes of this problem by looking at the terrorists themselves, the situation around them, and others' responses to them.

The Problem of Intuition

The results of these "everyday" research projects teaches us about human behaviour. We learn through experience what happens when we give someone bad news, that some people develop depression, and that aggressive behaviour occurs frequently in our society. We develop theories to explain all of these occurrences; however, it is important to remember that everyone's experiences are somewhat unique. My theory about why people suffer from depression may be completely different from yours, yet we both feel as though we are "right." The obvious problem here is that we cannot generalize from one person's experiences to people in general. We might both be wrong!

The problem with the way people collect and interpret data in their everyday lives is that they are not always particularly thorough or accurate. Often, when one explanation for an event seems right, we adopt that explanation as the truth even when other explanations are possible and potentially more accurate. Furthermore, we fall victim to confirmation bias; that is, we tend to seek information that confirms our beliefs regardless of the accuracy of those beliefs and discount any evidence to the contrary. Psychologists have found a variety of cognitive and motivational biases that frequently influence our perceptions and lead us to draw erroneous conclusions (Fiske & Taylor, 2007; Hsee & Hastie, 2006; Kahneman, 2011). Even feeling confident about our beliefs is not an

indicator of their accuracy. For example, eyewitnesses to violent crimes are often extremely confident in their identifications of the perpetrators of these crimes, but research finds that eyewitnesses are no less confident in their identifications when they are incorrect than when they are correct (Cutler & Wells, 2009; Wells & Hasel, 2008). In summary, accepting explanations without empirical evidence may lead us to faulty thinking and erroneous conclusions. Our faulty thinking is not limited to the present; it also occurs when we try to make sense of the past. We have a tendency to tell ourselves "I knew it all along" when making sense of past events; this is known as hindsight bias (Kahneman, 2011). Thus, one of the goals of psychology education is to make people become better thinkers, better consumers of ideas, and better at understanding how our own biases get in the way of true knowledge.

Why Scientists Rely on Empirical Methods

All scientists, whether they are physicists, chemists, biologists, sociologists, or psychologists, use empirical methods to study the topics that interest them. Empirical methods include the processes of collecting and organizing data and drawing conclusions about those data. The empirical methods used by scientists have developed over many years and provide a basis for collecting, analyzing, and interpreting data within a common framework in which information can be shared. We can label the scientific method as the set of assumptions, rules, and procedures that scientists use to conduct empirical research.

Although scientific research is an important method of studying human behaviour, not all questions can be answered using scientific approaches. Statements that cannot be objectively measured or objectively determined to be true or false are not within the domain of scientific inquiry. Scientists therefore draw a distinction between values and facts. Values are personal statements, such as "Abortion

should not be permitted in this country," "I will go to heaven when I die," or "It is important to study biology." Facts are objective statements determined to be accurate through empirical study. Examples are "The homicide rate in Canada has been generally declining over the past 45 years" or "Research demonstrates that individuals who are exposed to highly stressful situations over long periods of time develop more health problems than those who are not."

When we try to find new facts, we express our prediction about what we believe to be true in a hypothesis. An example of a hypothesis would be "People who eat fruits and vegetables daily have better health than people who never eat fruits and vegetables." For this to become fact, we must test this hypothesis in research and show the evidence that supports it. This is a testable hypothesis, because it would be possible to do the research. It is also falsifiable, meaning that if our prediction is wrong, and eating fruits and vegetables daily does not lead to better health, we will have the data to show us that we are wrong.

Ideas or values are not always testable or falsifiable: science can neither prove nor disprove them. For example, the famous Viennese neurologist Sigmund Freud, father of psychoanalysis, believed that the unconscious part of our mind is ultimately responsible when we experience anxiety, depression, and other negative emotions. He thought that emotional conflicts and adverse childhood experiences became lodged in the unconscious because consciously acknowledging them was threatening to our sense of wellbeing. This theory is built on a largely untestable idea: the existence of an unconscious. Given that, by definition, we cannot describe it, it is difficult to see how we could prove its existence or its role in our lives. That does not mean that the unconscious does not exist, but as we will see, we need to find a way to look for its existence using testable and falsifiable hypotheses.

Although scientists use research to help establish facts, the distinction between values or opinions and facts is not always clear-cut. Sometimes, statements that scientists consider to be factual

turn out to be partially or even entirely incorrect when researched further. Although scientific procedures do not necessarily guarantee that the answers to questions will be objective and unbiased, science is still the best method for drawing objective conclusions about the world around us. When old facts are discarded, they are replaced with new facts based on new and correct data. Although science is not perfect, the requirements of empiricism and objectivity result in a much greater chance of producing an accurate understanding of the world than what is available through other approaches.

Levels of Explanation

How we study a phenomenon can differ depending on the granularity of the data and observation. For example, we can study chemical bonds through an understanding of substances and their reactivity. Going into further detail, we can look at the molecular structure of each substance and even further down towards subatomic particles. However, what level you choose may often depend on your research question and the purpose of your inquiry. To take another example, scientists could use Einstein's theory of gravity and his equations to calculate trajectories for rockets sent into space. However, they can do just as well with Newton's equations simply because they take into account a smaller number of variables and provide sufficiently accurate outcomes for rocket launches. We can see such levels of analysis being illustrated in other fields, such as psychology, as well.

The study of psychology spans many different topics at many different levels of explanation. Lower levels of explanation are more closely tied to biological influences—such as genes, neurons, neurotransmitters, and hormones—whereas the middle levels of explanation refer to the abilities and characteristics of individual people, and the highest levels of explanation relate to social groups, organizations, and cultures (Cacioppo et al., 2000).

The same topic can be studied within psychology at different levels of explanation, as shown in the table below. For instance, the psychological disorder known as depression affects millions of people worldwide and is known to be caused by biological, social, and cultural factors. Studying and helping alleviate depression can be accomplished at low levels of explanation by investigating how chemicals in the brain influence the experience of depression. This approach has allowed psychologists to develop and prescribe drugs, such as Prozac, which may decrease depression in many individuals (Williams et al., 2009). At the middle levels of explanation,

psychological therapy is directed at helping individuals cope with negative life experiences that may cause depression. At the highest level, psychologists study differences in the prevalence of depression between men and women and across cultures. The occurrence of psychological disorders, including depression, is substantially higher for women than for men, and it is also higher in Western cultures, such as in Canada, the United States, and Europe, than in Eastern cultures, such as in India, China, and Japan (Chen et al., 2009; Seedat et al., 2009). These sex and cultural differences provide insight into the factors that cause depression. The study of depression in psychology helps remind us that no one level of explanation can explain everything. All levels of explanation, from biological to personal to cultural, are essential for a better understanding of human behaviour.

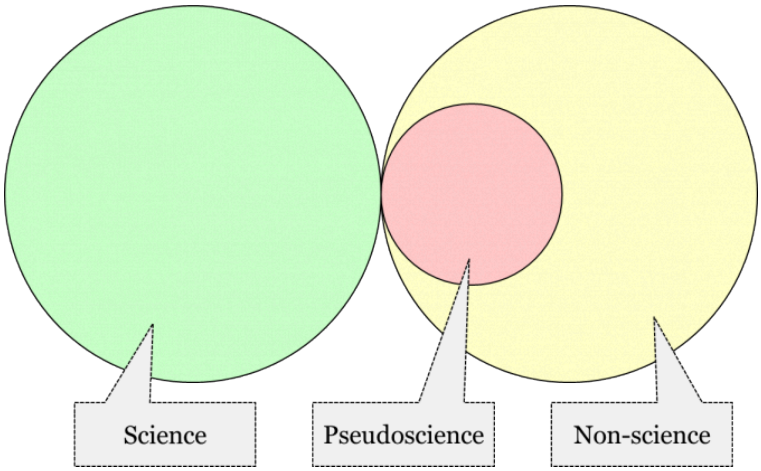
Jessica/Feb 20: Table caption missing and table number missing

Level of Explanation	Underlying Process	Examples
Lower	Biological	• Depression is, in part, genetically influenced. • Depression is influenced by the action of neurotransmitters in the brain.
Middle	Interpersonal	• People who are depressed may interpret the events that occur to them too negatively. • Psychotherapy can be used to help people talk about and combat depression.
Higher	Cultural and social	• Women experience more depression than do men. • The prevalence of depression varies across cultures and historical time periods.

Science and Non-science

What makes science what it is, or what differentiates it from other human endeavours? We have taken for granted that science is something different—something unique. But what actually makes science different? What makes it unique? In philosophy, this question has been called the demarcation problem. To demarcate something is to set its boundaries or limits, to draw a line between one thing and another. For instance, a white wooden fence demarcates my backyard from the backyard of my neighbour, and a border demarcates the end of one country's territory and the beginning of another's. The demarcation problem in the philosophy of science asks: What is the difference between science and non-science? In other words, what line—if any—separates science from non-science, and where exactly does science begin and non-science end?

Historically, many philosophers have sought to demarcate science from non-science. However, often, their specific focus has been on the demarcation between science and pseudoscience. Now, what is pseudoscience and how is it different from non-science in general? Pseudoscience is a very specific subspecies of non-science that masks itself as science. Consider, for instance, the champions of intelligent design, who essentially present their argument for the existence of God as a properly scientific theory that is purportedly based on scientific studies in fields such as molecular biology and evolutionary biology but incorporates both blatant and subtle misconceptions about evolutionary biology. Not only is the theory of intelligent design unscientific, but it is pseudoscientific, as it camouflages and presents itself as a legitimate science. In short, while not all non-science is pseudoscience, all pseudoscience is definitely non-science.



Jessica/Feb 20: *Image caption and attribution missing*

Pseudoscience Alert! Intelligent Design

"In crossing a heath, suppose I pitched my foot against a stone, and were asked how the stone came to be there; I might possibly answer, that, for anything I knew to the contrary, it had lain there forever: nor would it perhaps be very easy to show the absurdity of this answer. But suppose I had found a watch upon the ground, and it should be inquired how the watch happened to be in that place; I should hardly think of the answer I had before given, that for anything I knew, the watch might have always been there... There must have existed, at some time, and at some place or other, an artificer or artificers, who formed [the watch] for the purpose which

we find it actually to answer; who comprehended its construction, and designed its use. … Every indication of contrivance, every manifestation of design, which existed in the watch, exists in the works of nature; with the difference, on the side of nature, of being greater or more, and that in a degree which exceeds all computation!"

—William Paley, *Natural Theology* (1802)

This argument by William Paley from the 19th century is of one that has been popular among religious circles for centuries. However, this idea has emerged in recent years with new names such as creationism and intelligence design (ID). The idea is that given how complex life (or the universe) is, it must follow that there must be a complex designer to design it. More recently, the Discovery Institute (a think tank in the United States) has attempted to infiltrate science classrooms with books and propaganda advocating for this to be taught as a scientific fact. Their modern attempts have moved from Paley's watch and eye examples to mechanisms in bacterial flagella (but the argument is the same).

ID seeks to topple the methodological naturalism inherent in modern science. Methodological naturalism attempts to explain the universe and everything in it using natural laws and forces as opposed to supernatural entities. In challenging this idea, ID attempts to indoctrinate school children with Christian creation stories as scientific theories.

While pseudoscience is the most dangerous subspecies of non-science, philosophical discussions of the problem of demarcation aim to extract those features that make science what it is. Thus,

they concern the distinction between science and non-science in general, not only that between science and pseudoscience. So, our focus in this chapter is not only on pseudoscience exclusively, but on the general demarcation between science and non-science.

Practical Implications

As with most philosophical questions concerning science, this question has far-reaching practical implications. The question of demarcation is of great importance to policymaking, courts, healthcare, education, and journalism, as well as for the proper functioning of grant agencies. To appreciate the practical importance of the problem of demarcation, let us imagine what would happen if there was no way of telling science from non-science. Let us consider some of these practical implications in turn.

Suppose a certain epistemic community argues that we are facing a potential environmental disaster: say, an upcoming massive earthquake, an approaching asteroid, or slow but steady global warming. How seriously should we take such a claim? Naturally our reaction would depend on how trustworthy we think the position of this community is. We would probably not be very concerned if this was a claim championed exclusively by an unscientific—or worse, pseudoscientific—community. However, if the claim about looming disaster was accepted by a scientific community, it would likely have serious effect on our environmental policy and our decisions going forward. But this means that we need to have a way of telling what is science and what is not.

The ability to demarcate science from non-science and pseudoscience is equally important in courts, which customarily rely on the testimony of experts from different fields of science. Since litigating sides have a vested interest in the outcome of the litigation, they might be inclined towards using any available "evidence" in their favour, including "evidence" that has no scientific

foundation whatsoever. Thus, knowing what is science and what is not is very important for courts to function properly. Consider, for example, the ability to distinguish between claimed evidence obtained by psychic channelling and evidence obtained by analysing DNA found in blood at the scene of the crime.

The demarcation of science from non-science is also crucial for healthcare. It is an unfortunate fact that, in medicine, the promise of an easy profit often attracts those who are quick to offer "treatments" whose therapeutic efficacy has not been properly established. Such "treatments" can have health- and even life-threatening effects. Thus, any proper healthcare system should use only those treatments whose therapeutic efficacies have been scientifically established. But this assumes a clear understanding as to what is science and what merely masks itself as such.

A solid educational system is one of the hallmarks of a contemporary civilized society. It is commonly understood that we should not teach our children any pseudoscience; instead, we should build our curricula around knowledge accepted by our scientific community. For that reason, we do not think astrology, divination, or creation science have any place in school or university curricula. Of course, we sometimes discuss these subjects in history and philosophy of science courses, where they are studied as examples of non-science or of what was once considered scientific but is currently deemed unscientific. Importantly, however, we do not present them as accepted science. Therefore, as teachers, we must be able to tell pseudoscience from proper science.

In recent years, there have been several organized campaigns to portray pseudoscientific theories as bearing the same level of authority as the theories accepted by proper science. With the advent of social media, such as YouTube or Facebook, these campaigns become increasingly easy to orchestrate. Consider, for instance, the deniers of climate change or deniers of the efficacy of vaccination who have managed—through orchestrated journalism—to portray their claims as a legitimate stance in a scientific debate. Journalists should be properly educated to know

the difference between science and pseudoscience, for they otherwise risk hampering public opinion and dangerously influencing policymakers. Once again, this requires a philosophical understanding on how to demarcate science from non-science.

Finally, scientific grant agencies heavily relay on certain demarcation criteria when determining what types of research to fund and what types of research not to fund. For instance, these days we clearly would not fund an astrological project on the specific effect of, say, Jupiter's moons on a person's emotional makeup, while we would consider funding a psychological project on the effect of school-related stress on the emotional makeup of a student. Such decisions assume an ability to demarcate a scientific project from unscientific projects. In brief, the philosophical problem of demarcation between science and non-science is of great practical importance for a contemporary civilized society and its solution is a task of utmost urgency.

What are the Characteristics of a Scientific Theory?

Traditionally, the problem of demarcation has dealt mainly with determining whether certain theories are scientific or not. That is, in order to answer the more general question of distinguishing science and non-science, philosophers have focused on answering the more specific question of identifying features that distinguish scientific theories from unscientific theories. Thus, they have been concerned with the question: What are the characteristics of a scientific theory?

This more specific question treats the main distinction between science and non-science as a distinction between two different kinds of theories. Philosophers have therefore been trying to determine what features scientific theories have which unscientific theories lack. Consider for instance the following questions:

- Why is the theory of evolution scientific and creationism unscientific?
- Is the multiverse theory scientific?
- Are homeopathic theories pseudoscience?

Our contemporary scientific community answers questions like these on a regular basis, assessing theories and determining whether those theories fall within the limits of science or sit outside those limits. That is, the scientific community seems to have an implicit set of demarcation criteria that it employs to make these decisions. So, what are the demarcation criteria that scientists employ to evaluate whether a theory is scientific? First, let us look at our implicit expectations for what counts as a science and what does not. What are our current demarcation criteria? What criteria does the contemporary scientific community employ to determine which theories are scientific and which are not? Can we discover what they are and make them explicit? We can, but it will take a little bit of work. We will discuss a number of different characteristics of scientific theories and see whether those characteristics meet our contemporary implicit demarcation criteria. By considering each of these characteristics individually, one step at a time, we can hopefully refine our initially proposed criteria and build a clearer picture of what our implicit demarcation criteria actually are.

Note that, for the purposes of this exercise, we will focus on attempting to explicate our contemporary demarcation criteria for empirical science (as opposed to formal science). Empirical theories consist not merely of analytic propositions (i.e., definitions of terms and everything that follows from them) but also of synthetic propositions (i.e., claims about the world). This is true by definition: a theory is said to be empirical if it contains at least one synthetic proposition. Therefore, empirical theories are not true by definition: they can either be confirmed by our experiences or contradicted by them. So, propositions like "the Moon orbits the earth at an average distance of 384,400 km," or "a woodchuck could chuck 500 kg of wood per day," or "aliens created humanity and manufactured

the fossil record to deceive us" are all empirical theories because they could be confirmed or contradicted by experiments and observations. We will therefore aim to find out what our criteria are for determining whether an empirical theory is scientific or not.

First, let us appreciate that not all empirical theories are scientific. Consider the following example:

- **Theory A: You are currently in Kamloops, BC, Canada.**

That is right: I, the author, am making a claim about you, the reader. Right now. Theory A has all the hallmarks of an empirical theory: it is not an analytic proposition because it is not true by definition, and depending on your personal circumstances, it might be correct or incorrect. But it is not based on experience because I, the author, have no reason to think that you are, in fact, in Kamloops, BC: I have never seen you near Victoria Street, and there is no way you would choose reading your textbook over a day at Aberdeen Mall. Theory A is a genuine claim about the world, but it is a claim that is in a sense "cooked-up" and based on no experience whatsoever. Here are two other examples of empirical theories not based on experience:

- **Theory B: The ancient Romans moved their civilization to an underground location on the far side of the moon.**
- **Theory C: A planet 3 billion light years from Earth also has a company called Netflix.**

Therefore, we can safely conclude that not every empirical theory can be said to be scientific. If that is so, then what makes a particular empirical theory scientific? Let us start out by suggesting something simple.

- **Suggestion 1: An empirical theory is scientific if it is based on experience.**

This seems obvious, or maybe not even worth mentioning. After all, all empirical theories have to be based on experience, right?

Suggestion 1 is based on the fact that we do not want to consider empirical theories like A, B, and C to be scientific theories. Theories that we can come up with on a whim, grounded in no experience whatsoever, do not strike us as scientific. Rather, we expect that even simple scientific theories must be somehow grounded in our experience of the world.

This basic contemporary criterion that empirical theories be grounded in our experience has deep historical roots but was perhaps most famously attributed to British philosopher John Locke (1632–1704) in his text *An Essay Concerning Human Understanding*. In this work, Locke (1998) attempted to lay out the limits of human understanding, ultimately espousing a philosophical position known today as empiricism. Empiricism is the belief that all synthetic propositions (and consequently, all empirical theories) are justified by our sensory experiences of the world (i.e., by our experiments and observations). Empiricism stands against the position of apriorism (also often referred to as "rationalism")—another classical conception that was advocated for by the likes of René Descartes and Gottfried Wilhelm Leibniz. According to apriorists, there are at least some fundamental synthetic propositions that are knowable independently of experiments and observations, i.e., a priori (in philosophical discussions, a "priori" means "knowable independently of experience"). It is this idea of a priori synthetic propositions that apriorists accept and empiricists deny. Thus, the criterion that all physical, chemical, biological, sociological, and economical theories must be justified by experiments and observations only can be traced back to empiricism.

But is this basic criterion sufficient to properly demarcate scientific empirical theories from unscientific ones? If an empirical theory is based on experience, does that automatically make it scientific?

Perhaps the main problem with Suggestion 1 can best be illustrated with an example. Consider the contemporary opponents of vaccination, called "anti-vaxxers." Many anti-vaxxers today accept the following theory:

- **Theory D: Vaccinations are a major contributing cause of autism.**

This theory results from sorting through incredible amounts of medical literature and gathering patient testimonials. Theory D is clearly an empirical theory, and interestingly, it is also sense based on experience in some sense. As such, it seems to satisfy the criterion we came up with in Suggestion 1: Theory D is both empirical and is based on experience.

However, while being based on experience, Theory D also results from willingly ignoring some of the known data on that topic. A small study by Andrew Wakefield et al., published in *The Lancet* in 1998, became infamous around 2000–2002 when the UK media caught hold of it. In that article, the author hypothesized an alleged link between the measles vaccine and autism despite a small sample size of only 12 children. Theory D fails to take into account the sea of evidence suggesting both that no such link (between vaccines and autism) exists and that vaccines are essential to societal health.

In short, Suggestion 1 allows for theories that have "cherry-picked" their data to be considered scientific, since it allows scientific theories to be based on any arbitrarily selected experiences. This does not seem to agree with our implicit demarcation criteria. Theories like Theory D, while based on experience, are not generally considered to be scientific. As such, we need to refine the criterion from Suggestion 1 to see if we can avoid the problems illustrated by the anti-vaxxer Theory D. Consider the following alternative:

- **Suggestion 2: An empirical theory is considered scientific if it explains all the known facts of its domain.**

This new suggestion has several interesting features worth highlighting.

First, note that it requires a theory to explain all the known facts of its domain, and not only a selected ("cherry-picked") subset of the

known facts. By ensuring that a scientific empirical theory explains the "known facts," Suggestion 2 is clearly committed to being "based on experience." In this, it is similar to Suggestion 1. However, Suggestion 2 also stipulates that a theory must be able to explain all of the known facts of its domain to avoid the cherry-picking exemplified by the anti-vaxxer Theory D. As such, Suggestion 2 excludes theories that clearly cherry-pick their evidence and disqualifies such fabricated theories as unscientific. Therefore, theories that choose to ignore great swathes of relevant data, such as decades of research on the causes of autism, can be deemed unscientific by Suggestion 2.

Furthermore, Suggestion 2 explicitly talks about the facts within a certain domain. A domain is an area (field) of scientific study. For instance, life and living processes are the domain of biology, whereas the Earth's crust, processes in it, and its history are the domain of geology. By specifying that an empirical theory has to explain the known facts of its domain, Suggestion 2 simply imposes more realistic expectations: it does not expect theories to explain all the known facts from all fields of inquiry. In other words, it does not stipulate that, in order to be scientific, a theory should explain everything. For instance, if an empirical theory is about the causes of autism (like Theory D), then the theory should merely account for the known facts regarding the causes of autism, not the known facts concerning black holes, the evolution of species, or inflation.

Does Suggestion 2 hold water? Can we say that it correctly explicates the criteria of demarcation currently employed in empirical science? The short answer is *not quite*.

When we look at general empirical theories that we unproblematically consider scientific, like the theory of general relativity or the theory of evolution by natural selection, we notice that even they may fail to meet the stringent requirements of Suggestion 2. Indeed, do our best scientific theories explain all the known facts of their respective domains? Can we reasonably claim that the current biological theories explain all the known biological

facts? Similarly, can we say that our accepted physical theories explain all the known physical phenomena?

It is easy to see that even our best accepted scientific theories today cannot account for absolutely every known piece of data in their respective domains. It is a known historical fact that scientific theories rarely succeed in explaining all the known phenomena of their domain. In that sense, our currently accepted theories are no exception.

Take, for example, the theory of evolution by natural selection. Our contemporary scientific community clearly considers evolutionary theory to be a proper scientific theory. However, it is generally accepted that evolution itself is a very slow process. For the first 3.5 billion years of life's history on Earth, organisms evolved from simple single-celled bacterium-like organisms to simple multicellular organisms like sponges. About 500 million years ago, however, there was a major—relatively sudden (on a geological timescale of millions of years)—diversification of life on Earth that scientists call the Cambrian explosion, wherein we see the beginnings of many of the animal life forms we are familiar with today, such as arthropods, molluscs, chordates, etc. Nowadays, biologists accept both the existence of the Cambrian explosion and the theory of evolution by natural selection. Nevertheless, the theory does not currently explain the phenomenon of the Cambrian explosion. In short, what we are dealing with here is a well-known fact in the domain of biology, which our accepted biological theory does not explain.

What this example demonstrates is that scientific theories do not always explain all the known facts of its domain. Thus, if we were to apply Suggestion 2 in actual scientific practice, we would have to exclude virtually all of the currently accepted scientific empirical theories. This means that Suggestion 2 cannot possibly be the correct explication of our current implicit demarcation criterion. So, let us make a minor adjustment:

- **Suggestion 3: An empirical theory is scientific if it explains,**

by and large, the known facts of its domain.

Like Suggestion 2, this formulation of our contemporary demarcation criterion also ensures that scientific theories are based on experience and not cherry-picked data. However, it introduces an important clause—"by and large"—and thus clarifies that an empirical theory simply has to account for the great majority of the known facts of its domain. Note that this new clause is not quantitative: it does not stipulate what percentage of the known facts must be explained. The clause is qualitative, as it requires a theory to explain virtually all but not necessarily all the known facts of its domain. This simple adjustment accomplishes an important task of imposing more realistic requirements. Specifically, unlike Suggestion 2, Suggestion 3 avoids excluding theories like the theory of evolution from science.

How well does Suggestion 3 square with the actual practice of science today? Does it come close to correctly explicating the actual demarcation criteria employed nowadays in empirical science?

To the best of our knowledge, Suggestion 3 seems to be a necessary condition for an empirical theory to be considered scientific today. That is, any theory that the contemporary scientific community deems scientific must explain, by and large, the known facts of its domain. Yet while the requirement to explain most facts of a certain domain seems to be a necessary condition for being considered a scientific theory today, it is not a sufficient condition. Indeed, while every scientific theory today seems to satisfy the criterion outlined in Suggestion 3, that criterion on its own does not seem to be sufficient to demarcate scientific theories from unscientific theories. The key problem here is that some famous unscientific theories also manage to meet this criterion.

Take, for example, the theory of astrology. Among many other things, astrology claims that processes on the Earth are at least partially due to the influence of the stars and planets. Specifically, astrology suggests that a person's personality and traits depend crucially on the specific arrangement of planets at the moment of

their birth. As the existence of such a connection is far from trivial, astrology can be said to contain synthetic propositions, by virtue of which it can be considered an empirical theory. Now, it is clear that astrology is notoriously successful at explaining the known facts of its domain. If the Sun was in the constellation of Taurus at the moment of a person's birth, and this person happened to be even somewhat persistent, then this would be in perfect accord with what astrology says about Tauruses. But even if this person was not persistent at all, a trained astrologer could still explain the person's personality traits by referring to the subtle influences of other celestial bodies. No matter how much a person's personality diverges from the description of their "sign," astrology somehow always finds a way to explain it. As such, astrology can be considered an empirical theory that explains, by and large, the known facts of its domain, as per Suggestion 3.

This means that while Suggestion 3 seems to faithfully explicate a necessary part of our contemporary demarcation criteria, there must be at least another necessary condition. It should be a condition that the theory of evolution and other scientific theories satisfy, while astrology and other unscientific theories do not. What can this additional condition be?

One idea that comes to mind is that of testability. Indeed, it seems customary in contemporary empirical science to expect a theory to be testable. Thus, it seems that in addition to explaining, by and large, the known facts of their domains, scientific empirical theories are also expected to be testable, at least in principle.

It is important to appreciate that what is important here is not whether we as a scientific community currently have the technical means and financial resources to test the theory. Instead, what is important is whether a theory is testable in principle. In other words, we seem to require that there be a conceivable way of testing a theory, regardless of whether it is or is not possible to conduct that testing in practice. Suppose there is a theory that makes some bold claims about the structure and mechanism of a certain subatomic process. Suppose also that the only way of testing the

theory that we could think of is by constructing a gigantic particle accelerator the size of the solar system. Clearly, we are not in a position to actually construct such an enormous accelerator for obvious technological and financial reasons. Such issues actually arise in string theory, a pursued attempt to combine quantum mechanics with general relativity theory into a single consistent theory. They are a matter of strong controversy among physicists and philosophers. What seems to matter to scientists is merely the ability of a theory to be tested in principle. In other words, even if we have no way of testing a theory currently, we should at least be able to conceive of a means of testing it. If there is no conceivable way of comparing the predictions of a theory to the results of experiments or observations, then it would be considered untestable, and therefore unscientific.

But what exactly does the requirement of testability imply? How should testability itself be understood? In the philosophy of science, there have been many attempts to clarify the notion of testability. Two opposing notions of testability are particularly notable—verifiability and falsifiability. Let us consider these in turn.

Among others, Rudolph Carnap suggested that an empirical theory is scientific if it has the possibility of being verified in experiments and observations. For Carnap, a theory was considered verified if predictions of the theory could be confirmed through experience. Take the simple empirical theory:

- **Theory E: The light in my refrigerator turns off when I close the door.**

If I set up a video camera inside the fridge so that I could see that the light does, indeed, turn off whenever I close the door, Carnap would consider the theory to be verified by my experiment. According to Carnap, every scientific theory is like this: we can, in principle, find a way to test and confirm its predictions. This position is called verificationism. According to verificationism, an empirical theory

is scientific if it is possible to confirm (verify) the theory through experiments and observations.

Alternatively, Karl Popper (2002) suggested that an empirical theory is scientific if it has the possibility of being falsified by experiments and observations. Whereas Carnap focused on the ability of theories to become verified by experience, Popper held that what truly makes a theory scientific is its potential ability to be disconfirmed by experiments and observation. Science, according to Popper, is all about bold conjectures that are tested and tentatively accepted until they are falsified by counterexamples. The ability to withstand any conceivable test, for Popper, is not a virtue but a vice that characterizes all unscientific theories. What makes Theory E scientific, for Popper, is the fact that we can imagine the possibility that what I see on the video camera when I close my fridge might not match my theory. If the light, in fact, does not turn off when I close the door a few times, then Theory E would be considered falsified. What matters here is not whether a theory has or has not actually been falsified, or even if we have the technical means to falsify the theory, but whether its falsification is conceivable, i.e., whether there can, in principle, be an observational outcome that would falsify the theory. According to falsificationism, a theory is scientific if it can conceivably be shown to conflict with the results of experiments and observations.

Falsifiability and verifiability are two distinct interpretations of what it means to be testable. While both verifiability and falsifiability have their issues, the requirement of falsifiability seems to be closer to the current expectations of empirical scientists. Let us look briefly at the theory of young-Earth creationism to illustrate why.

Young-Earth creationists hold that the Earth, and all life on it, was directly created by God less than 10,000 years ago. While fossils and the layers of the Earth's crust appear to be millions or billions of years old according to today's accepted scientific theories, young-Earth creationists believe that they are not. In particular, young-Earth creationists subscribe to:

- **Theory F: Fossils and rocks were created by God within the last 10,000 years but were made by God to appear like they are over 10,000 years old.**

Now, is this theory testable? The answer depends on whether we understand testability as verifiability or falsifiability. Let us first see if Theory F is verifiable.

By the standards of verificationism, Theory F is verifiable, since it can be tested and confirmed by the data from experiments and/or observations. This is so because any fossil, rock, or core sample that is measured to be older than 10,000 years will actually confirm the theory, since Theory F states that such objects were created to seem that way. Every ancient object further confirms Theory F, and—from the perspective of verificationism—these confirmations would be evidence of the theory's testability. Therefore, if we were to apply the requirement of verifiability, young-Earth creationism would likely turn out scientific.

In contrast, by the standards of falsificationism, Theory F is not falsifiable; we can try and test it as much as we please, but we can never show that Theory F contradicts the results of experiments and observations. This is so because even if we were to find a trillion-year-old rock, it would not in any way contradict Theory F, since proponents of Theory F would simply respond that God made the rock to seem one trillion years old. Theory F is formulated in such a way that no new data (i.e., no new evidence) could ever possibly contradict it. As such, from the perspective of falsificationism, Theory F is untestable and thus unscientific.

Understanding a theory's testability as its falsifiability seems to be the best way to explicate this second condition of our contemporary demarcation criteria: for the contemporary scientific community, to say that a theory is testable is to say that it is falsifiable, i.e., that it can, in principle, contradict the results of experiments and observations. With this understanding of testability clarified, it seems we have our second necessary condition for a theory to be considered scientific. To sum up, in our contemporary empirical

science, we seem to consider a theory scientific if it explains, by and large, the known facts of its domain, and it is testable (falsifiable), at least in principle:

Contemporary Demarcation Criteria

An empirical theory is scientific if it explains, by and large, the known facts of domain and it is testable (falsifiable), at least in principle.

We began this exercise as an attempt to explicate our contemporary criteria for demarcation, and we have done a lot of work to distil the contemporary demarcation criteria above. It is important to note that this is merely our attempt at explicating the contemporary demarcation criteria. Just as with any other attempt to explicate a community's method, our attempt may or may not be successful. Since we were trying to make explicit those implicit criteria employed to demarcate science from non-science, even this two-part criterion might still need to be refined further. It is quite possible that the actual demarcation criteria employed by empirical scientists are much more nuanced and contain many additional clauses and sub-clauses. That being said, we can take our explication as an acceptable first approximation of the contemporary demarcation criteria employed in empirical science.

This brings us to one of the central questions of this chapter. Suppose, for the sake of argument, that the contemporary demarcation criteria are along the lines of our explication above, i.e., that our contemporary empirical science indeed expects scientific theories to explain by and large, the known facts of its domain and be, in principle, falsifiable. Now, can we legitimately claim that these same demarcation criteria have been employed in all time periods? That is, could these criteria be the universal and transhistorical

criteria of demarcation between scientific and unscientific theories? More generally:

Are there universal and transhistorical criteria for demarcating scientific theories from unscientific theories?

The short answer to this question is no. There are both theoretical and historical reasons to believe that the criteria that scientists employ to demarcate scientific from unscientific theories are neither fixed nor universal. Both the history of science and the laws of scientific change suggest that the criteria of demarcation can differ drastically across time periods and fields of inquiry. Let us consider the historical and theoretical reasons in turn.

For our theoretical reason, let us look at the laws of scientific change. Recall the third law of scientific change, the laws of method employment, which states that newly employed methods are the deductive consequences of some subset of other accepted theories and employed methods. As such, when theories change, methods change with them. This holds equally for the criteria of acceptance, criteria of compatibility, and criteria of demarcation. Indeed, since the demarcation criteria are part of the method, demarcation criteria change in the same way that all other criteria do: they become employed when they follow deductively from accepted theories and other employed methods. As such, the demarcation criteria are not immune to change, and therefore our contemporary demarcation criteria—whatever they are—cannot be universal or unchangeable.

This is also confirmed by historical examples. The historical reason to believe that our contemporary demarcation criteria are neither universal nor unchangeable is that there have been other demarcation criteria employed in the past. Consider, for instance, the criteria of demarcation employed by many Aristotelian-Medieval communities. One of the essential elements of the Aristotelian-Medieval worldview was the idea that all things not crafted by humans have a nature, an indispensable quality that makes a thing what it is. It was also accepted that an experienced person can grasp this nature through intuition schooled by

experience. The Aristotelian-Medieval method of intuition was a deductive consequence of these two accepted ideas. According to their acceptance criteria, a theory was expected to successfully grasp the nature of a thing in order to become accepted. In their demarcation criteria, they stipulated that a theory should at least attempt to grasp the nature of a thing under study, regardless of whether it actually succeeded in doing so. Thus, we can explicate the Aristotelian-Medieval demarcation criterion as :

Aristotelian Demarcation Criteria

An empirical theory is scientific if it attempts to uncover the nature of a thing.

Thus, both natural philosophy and natural history were thought to be scientific: while natural philosophy was considered scientific because it attempted to uncover the nature of physical reality, natural history was scientific for attempting to uncover the nature of each creature in the world. Mechanics, however, was not considered scientific precisely because it dealt with things crafted by humans. As opposed to natural things, artificial things were thought to have no intrinsic nature but were created by a craftsman for the sake of something else. Clocks, for instance, do not exist for their own sake, but for the sake of timekeeping. Similarly, ships do not have any nature but are built to navigate people from place to place. Thus, according to the Aristotelians, the study of these artefacts (mechanics) is not scientific, since there is no nature for it to grasp.

It should be clear by now that the same theory could be considered scientific in one worldview and unscientific in a different worldview depending on the respective demarcation criteria employed in the two views. For instance, astrology satisfied

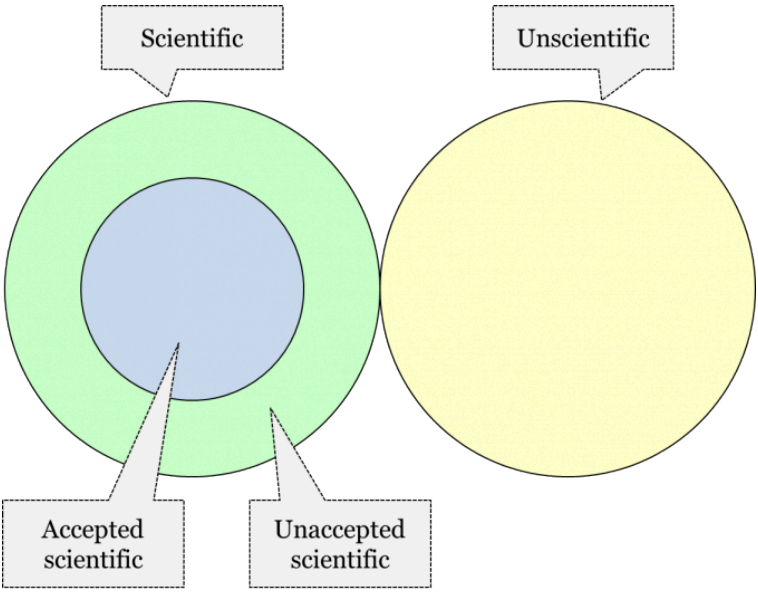
the Aristotelian-Medieval demarcation criteria, as it clearly attempted to grasp the nature of celestial bodies by studying their effects on the terrestrial realm. It was therefore considered scientific. As we know, astrology is not currently considered scientific, since it does not satisfy our current demarcation criteria. What this tells us is that demarcation criteria change over time.

Not only do they change over time, but they can also differ from one field of inquiry to another. For instance, while some fields seem to take the requirement of falsifiability seriously, other fields consider the very notion of empirical falsification to be problematic. This applies not only to formal sciences, such as logic and mathematics, but also to some fields in the social sciences and humanities.

In short, we have both theoretical and historical reasons to believe that there can be no universal demarcation criteria. Because demarcation criteria are specific to a certain time period, theories are appraised by different scientific communities at different periods of history using different criteria, and it follows that their appraisals of whether theories are scientific or not may differ.

Scientific vs. Unscientific and Accepted vs. Unaccepted

Before we proceed, it is important to restate that demarcation criteria and acceptance criteria are not the same thing, as they play different roles. While demarcation criteria are employed to determine whether a theory is scientific or not, acceptance criteria are employed to determine whether a theory ought to be accepted as the best available description of its object. Importantly, it is therefore possible for a community to consider a theory to be scientific and to nevertheless leave the theory unaccepted. Here is a Venn diagram illustrating the relations between the categories of unscientific, scientific, accepted, and unaccepted:



Jessica/Feb 20: Image caption and attribution missing

A few examples will help clarify the distinction. General relativity is considered to be both scientific and accepted, because it passed the strong test of predicting the degree to which starlight would be bent by the gravitational field of the sun, and other subsequent tests. In contrast, string theory is considered by the contemporary scientific community as a scientific theory, but it is not yet accepted as the best available physical theory. Alchemy had a status similar to that of string theory in the Aristotelian-Medieval world view. The Aristotelian-Medieval community never accepted alchemy but considered it to be a legitimate science.

Media Attributions

Jessica/ Feb 20: Figures and Information missing

Science Versus Pseudoscience

Pseudoscience refers to activities and beliefs that are claimed to be scientific by their proponents—and may appear to be scientific at first glance—but are not. Consider the theory of biorhythms (not to be confused with sleep cycles or circadian rhythms that do have a scientific basis). The idea is that people's physical, intellectual, and emotional abilities run in cycles that begin when they are born and continue until they die. Allegedly, the physical cycle has a period of 23 days, the intellectual cycle a period of 33 days, and the emotional cycle a period of 28 days. So, for example, if you had the option of when to schedule an exam, you would want to schedule it for a time when your intellectual cycle will be at a high point. The theory of biorhythms has been around for more than 100 years, and you can find numerous popular books and websites about biorhythms, often containing impressive and scientific-sounding terms like sinusoidal wave and bioelectricity. The problem with biorhythms, however, is that scientific evidence indicates they do not exist (Hines, 1998).

A set of beliefs or activities can be said to be pseudoscientific if (a) its adherents claim or imply that it is scientific but (b) it lacks one or more of the three features of science. For instance, it might lack systematic empiricism. Either there is no relevant scientific research or, as in the case of biorhythms, there is relevant scientific research, but it is ignored. It might also lack public knowledge. People who promote these beliefs or activities might claim to have conducted scientific research but never publish that research in a way that allows others to evaluate it.

A set of beliefs and activities might also be pseudoscientific because it does not address empirical questions. The philosopher Karl Popper (2002) was especially concerned with this idea. Specifically, he argued that any scientific claim must be expressed in such a way that there are observations that would—if they were made—count as evidence against the claim. In other words,

scientific claims must be falsifiable. The claim that women talk more than men is falsifiable because systematic observations could reveal either that they do talk more than men or that they do not. As an example of an unfalsifiable claim, consider that many people who believe in extrasensory perception (ESP) and other psychic powers claim that such powers can disappear when they are observed too closely. This makes it so that no possible observation would count as evidence against ESP. If a careful test of a self-proclaimed psychic showed that she predicted the future at better-than-chance levels, this would be consistent with the claim that she had psychic powers. But if she failed to predict the future at better-than-chance levels, this would also be consistent with the claim because her powers can supposedly disappear when they are observed too closely.

Why should we concern ourselves with pseudoscience? There are at least three reasons. One is that learning about pseudoscience helps bring the fundamental features of science—and their importance—into sharper focus. A second is that biorhythms, psychic powers, astrology, and many other pseudoscientific beliefs are widely held and promoted on the internet, on television, and in books and magazines. Far from being harmless, the promotion of these beliefs often results in great personal toll as, for example, believers in pseudoscience opt for "treatments" such as homeopathy for serious medical conditions instead of empirically supported treatments. Learning what makes them pseudoscientific can help us identify and evaluate such beliefs and practices when we encounter them. A third reason is that many pseudosciences claim to explain some aspect of human behaviour and mental processes, including biorhythms, astrology, graphology (handwriting analysis), and magnet therapy for pain control. It is important for students of psychology to clearly distinguish their own field from this "pseudo psychology."

Resources for Thinking

The Skeptic's Dictionary

An excellent source for information on pseudoscience is *The Skeptic's Dictionary* (http://www.skepdic.com). Among the pseudoscientific beliefs and practices, you can learn about are the following:

- **Cryptozoology:** The study of "hidden" creatures like Bigfoot, the Loch Ness monster, and the chupacabra.
- **Pseudoscientific psychotherapies:** Past-life regression, rebirthing therapy, and bioscream therapy, among others.
- **Homeopathy:** The treatment of medical conditions using natural substances that have been diluted, sometimes to the point of no longer being present.
- **Pyramidology:** Odd theories about the origin and function of the Egyptian pyramids (e.g., that they were built by extraterrestrials) and the idea that pyramids, in general, have healing and other special powers.

Another excellent online resource is *Neurobonkers* (http://neurobonkers.com), which regularly posts articles that investigate claims that pertain specifically to psychological science.

Psychology and Pseudoscience

It is important to understand both what psychology is and what it is not. Psychology is a science because it uses and relies on the scientific method as a way to acquire knowledge. Science is an open activity, meaning that results are shared and published. Many scientists conduct research on the same topic, although perhaps from slightly different angles. You can think of scientific knowledge as a snowball—the more knowledge we have, the bigger the snowball, and if it carries on rolling, scientists carry on conducting research at an issue from multiple angles and perspectives. Thus, science is essentially a collaborative process, with all aspects of the process and results shared with the wider community for consideration and analysis. In science, your conclusions have to be based on empirical evidence, and the merits of the conclusions will have to be judged by peers before they are published. This process is called peer review, and it ensures that what gets published has been reviewed for its adherence to scientific methods and for the accuracy of its conclusions.

Occasionally, we may read a news story or see an advertisement claiming to have information that will change our lives. Such stories often allude to evidence but fail to show where to find it, or they rely on anecdotal, rather than empirical, evidence. They sometimes use celebrity endorsements and claim that if we do not act immediately to buy their information or product, we will lose out. Typically, these claims are targeted at significant concerns, such as weight loss, unhappiness, unspecified pain, and the like—often the very areas that psychologists are attempting to learn more about. Pseudoscience is the term often used to describe such claims; equivalent terms are "bad science" and "junk science." Fortunately, we can use the critical thinking processes outlined above to evaluate the veracity of claims that seem too good to be true.

Summary

Key Takeaways

- All scientists rely on the scientific method.
- Science is an open activity, meaning that results are shared and published.
- Everyday intuitions are often used by people to interpret the world. However, they usually lead to errors in thinking.
- The empirical methods used by scientists have developed over many years and provide a basis for collecting, analyzing, and interpreting data within a common framework in which information can be shared.
- Sometimes, statements that scientists consider to be factual turn out to be partially or even entirely incorrect when researched further.
- Although scientific procedures do not necessarily guarantee that the answers to questions will be objective and unbiased, science is still the best method for drawing objective conclusions about the world around us.
- Pseudoscience refers to activities and beliefs that are claimed to be scientific by their proponents—and may appear to be scientific at first glance—but are not.

Acknowledgements

Parts of this chapter were adapted from the following Open Education Resources:

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.
- "1.1 Psychology as a Science" in *Psychology – 1st Canadian Edition* by Walters (2020), via Thompson Rivers University, is used under a CC BY-NC-SA 4.0 license.
- "2. Absolute Knowledge" in *Introduction to History and Philosophy of Science* by Hakob et al. (n.d.), via Open Library (eCampusOntario), is used under a CC BY 4.0 license.

References

Cacioppo, J. T., Berntson, G. G., Sheridan, J. F., & McClintock, M. K. (2000). Multilevel integrative analyses of human behavior: Social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin*, 126(6), 829–843. https://psycnet.apa.org/doi/10.1037/0033-2909.126.6.829

Chen, P.-Y., Wang, S.-C., Poland, R. E., & Lin, K.-M. (2009). Biological variations in depression and anxiety between East and West. *CNS Neuroscience & Therapeutics*, 15(3), 283–294. https://doi.org/10.1111/j.1755-5949.2009.00093.x

Cutler, B. L., & Wells, G. L. (2009). Expert testimony regarding eyewitness identification. In J. L. Skeem, S. O. Lilienfeld, & K. S. Douglas (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 100–123). Guilford Press.

Fiske, S. T., & Taylor, S. E. (2007). *Social cognition: From brains to culture*. McGraw-Hill.

Hines, T. M. (1998). Comprehensive review of biorhythm theory. *Psychological Reports*, 83(1), 19–64. https://doi.org/10.2466/pr0.1998.83.1.19

Hsee, C. K., & Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences*, 10(1), 31–37. https://doi.org/10.1016/j.tics.2005.11.007

Johnson, D. J., Cheung, F., & Donnellan, M. B. (2013). Does cleanliness influence moral judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology*, 45(3), 209–215. https://doi.org/10.1027/1864-9335/a000186

Kahneman, D. (2011). *Thinking, fast and slow*. Random House.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, S., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . Nosek, B. A. (2013). Investigating variation in replicability: A "many labs"

replication project. *Social Psychology*, 45(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Locke, J. (1998). *An essay concerning human understanding* (R. Woolhouse, Ed.). Penguin Classics.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Prentice Hall.

Paley, W. (1802). *Natural theology: Or, evidences of the existence and attributes of the deity; Collected from the appearances of nature*. R. Faulder.

Popper, K. R. (2002). *Conjectures and refutations: The growth of scientific knowledge*. Routledge.

Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, 19(12), 1219–1222. https://doi.org/10.1111/j.1467-9280.2008.02227.x

Seedat, S., Scott, K. M., Angermeyer, M. C., Berglund, P., Bromet, E. J., Brugha, T. S., & Kessler, R. C. (2009). Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. *Archives of General Psychiatry*, 66(7), 785–795. https://doi.org/10.1001/archgenpsychiatry.2009.36

Stanovich, K. E. (2010). *How to think straight about psychology* (9th ed.). Allyn & Bacon.

Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., Berelowitz, Dhillon, A. P., Thomson, M. A., Harvey, P., Valentine, A., Davies, S. E., & Walker-Smith, J. A. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, 351(9103). P637–P641. https://doi.org/10.1016/S0140-6736(97)11096-0

Wells, G. L., & Hasel, L. E. (2008). Eyewitness identification: Issues in common knowledge and generalization. In E. Borgida & S. T. Fiske (Eds.), *Beyond common sense: Psychological science in the courtroom* (pp. 159–176). Blackwell.

Williams, N., Simpson, A. N., Simpson, K., & Nahas, Z. (2009). Relapse rates with long-term antidepressant drug therapy: A meta-

analysis. *Human Psychopharmacology: Clinical and Experimental*, 24(5), 401–408. https://doi.org/10.1002/hup.1033

IV. INTUITION PUMPS

Learning Objectives

- Describe the eight processes used in critical thinking
- Differentiate between formal and informal fallacies
- Identify some common logical fallacies

Critical Thinking

—*Mahabharata* 3:314

Logic alone is insufficient. Neither revealed texts nor the holy seers can guide you. It is as if the true nature of *dharma* is hidden within a cave. The path may only be discerned by following the advice of great people.

Rational, objective thinking is a hallmark of science. Scientists need to be able to critically think about a problem or issue from a variety of perspectives, but it is not just scientists who need to be good thinkers. Critical thinking skills allow you to be a good consumer of ideas. Before deciding to buy a car, most people would spend some time evaluating what they wanted to buy, how much money they have to spend, what kind of car has a good safety record, and so on. The same thinking processes can be applied to ideas. These critical thinking processes are not necessarily intuitive. The good thing is that we can be taught what they are and how to use them. Thus, learning to think critically is a skill we can acquire with practise. Wade (1995) outlined eight processes that are used in critical thinking:

1. **Ask questions and be willing to wonder.**

Curiosity precedes all scientific discoveries and is the basis for acquiring knowledge. For example, suppose you are interested in whether physical exercise is important for mood. Questions you might be wondering about could include:

- Does getting exercise change your mood?

- Do people who work out daily feel happier than people who don't get any exercise?
- Does the kind of exercise matter?
- How much exercise makes a difference?
- How do you measure mood anyway?

2. **Define the problem.**

We need to think about exactly what we want to know about the connection between exercise and mood. There are many ways to define exercise: it might mean walking every day to some people or lifting weights or doing yoga to others. Similarly, some people might interpret mood to be something fairly fleeting, while other people might think mood refers to clinical depression. We need to define the issue or question we are interested in.

3. **Examine the evidence.**

Empirical evidence is the kind of support that critical thinkers seek. While you may be aware that a friend or relative felt better when they started running every day, that kind of evidence is anecdotal—it relates to one person, and we do not know if it would apply to other people. To look for evidence, we should turn to properly conducted studies. In psychology, these are most easily found in a searchable database called PsycINFO, which is available through university libraries. PsycINFO contains a vast index of all of the scholarly work published in psychology. Users can often download the original research articles straight from the PsycINFO website.

4. **Analyze assumptions and biases**

Whenever we are reasoning about an idea, we are bound to begin with certain assumptions. For example, we might assume that exercise is good for mood because we usually assume that there is no downside to exercise. It helps if we can identify how we feel

or think about an idea. All people are prone to some degree of confirmation bias—which is the tendency to look for evidence that supports your belief while, at the same time, discounting any that disproves it. This type of bias might be reflected in the social media accounts we follow—we follow people who think like us and do not follow people who might have opposing—yet possibly valid—points of view.

5. **Avoid emotional reasoning**

This process is related to the previous one. It is hard to think about anything in a completely objective manner. Having a vested interest in an issue, or personal knowledge about it, often creates an emotional bias that we may not even be aware of. Feeling strongly about something does not make us think rationally; in fact, it can be a barrier to rational thinking. Consider any issue you feel strongly about. How easy is it to separate your emotions from your objectivity?

6. **Avoid oversimplification**

Simplicity is comfortable, but it may not be accurate. We often strive for simple explanations for events because we do not have access to all of the information we need to fully understand the issue. This process relates to the need to ask questions. We should be asking ourselves "What do I not know about this issue?" Sometimes, issues are so complex that we can only address one little part. For example, many things are likely to affect mood; while we might be able to understand the connection to some types of physical exercise, we are not addressing any of the myriad of other social, cognitive, and biological factors that may be important.

7. **Consider other interpretations**

Whenever you hear a news story telling you that something is good for you, it is wise to dig a little deeper. For example, many news

stories report on research concerning the effects of alcohol. They may report that small amounts of alcohol have some positive health effects, that abstaining completely from alcohol is not good for you, and so on. A critical thinker would want to know more about how those studies were done, and they might suggest that perhaps moderate social drinkers differ from abstainers in a variety of lifestyle habits. Perhaps there are other interpretations for the link between alcohol consumption and health.

8. **Tolerate uncertainty**

Uncertainty is uncomfortable. We want to know why things happen for good reasons. We are always trying to make sense of the world, and we look for explanations. However, sometimes, things are complicated and uncertain, or we do not have an explanation for it yet. Sometimes, we just have to accept that we do not have a full picture of why something happens or what causes what yet. We need to remain open to more information. It is helpful to be able to point out what we do not know, as well as what we do.

Pseudoscience Alert: Your Handwriting Does Not Reveal Your Character

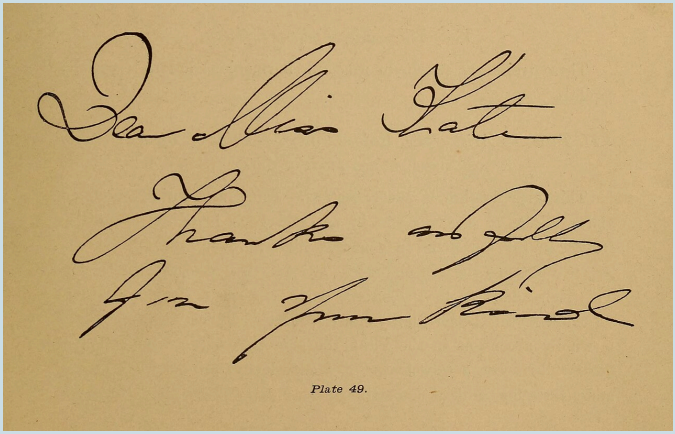


Figure 4.1 *A piece of handwriting used in graphological analysis, supposedly showing traits of "frivolity" and "triviality" in the writer Jessica/Feb 20: License of image missing*

One of the common beliefs that people have about personality is that it is revealed in one's handwriting. Graphology is a pseudoscience that purports to show that aspects of our handwriting reflect aspects of our character. *Reader's Digest* claims to explain the meaning of handwriting characteristics, such as spacing, size of characters, and how you cross your t's (LaBianca & Gibson, n.d.). "If you dot your i's high on the page, you likely have an active imagination, according to handwriting analysis experts. A closely dotted i is the mark of an organized and detail-oriented mind. If you dot your i's to the left, you might be a procrastinator, and if you dot your i's with a circle, you likely have playful and childlike qualities" (LaBianca & Gibson, n.d., "How do you dot your i's?").

Graphology has a long history (e.g., Münsterberg, 1915;

Downey, 1919) and is related to many other attempts to explain people's character based on aspects of physical appearance, such as the shape of one's face, size of one's hands, or bumps on one's head (i.e., phrenology). People have always been interested in personality and how to measure, describe, and explain it.

Unfortunately, this attempt to understand people's personality or career prospects by reading their handwriting has no empirical evidence to support it (Dazzi & Pedrabissi, 2009; Lilienfeld et al., 2010). Even though graphology was debunked several decades ago, the *Reader's Digest* article on their website in 2020 shows that the belief persists.

Media Attributions

Jessica/Feb 20: Information missing

Logical fallacies

A fallacy is simply a mistake in reasoning. Some fallacies are formal, while others are informal. Sometimes, we can define validity formally and thus determine whether an argument is valid or invalid without even having to know or understand what the argument is about. In other words, we could define certain valid rules of inference. These inference patterns are valid in virtue of their form, not their content. That is, any argument that has the same form as a valid argument will automatically be valid. A formal fallacy is simply an argument whose form is invalid. Thus, any argument that has that form will automatically be invalid, regardless of the meaning of the sentences. In contrast, informal fallacies are those that cannot be identified without understanding the concepts involved in the argument.

Formal fallacies

Denying the Antecedent

The *denying the antecedent* fallacy would have the following form:

1. If P, then Q.
2. Not P.
3. Therefore, not Q.

Consider the following argument:

1. If *Kant* was a *deontologist*, then he was a *non-consequentialist*.
2. *Kant* was not a *deontologist*.
3. Therefore, *Kant* was a not a *non-consequentialist*.

We know that this argument is invalid even if we do not know what "Kant" or "deontologist" or "non-consequentialist" means. ("Kant" was a famous German philosopher from the early 1800s, whereas "deontology" and "non-consequentialist" are terms that come from ethical theory.) It is mark of a formal fallacy that we can identify it even if we do not really understand the meanings of the sentences in the argument.

Affirming the Consequent

The *affirming the consequent* fallacy has the following form:

1. If P, then Q.
2. Q.
3. Therefore, P.

Here is an argument that uses silly, made-up words from Lewis Carroll's "Jabberwocky." See if you can determine whether the argument's form is valid or invalid:

1. If *toves* are *brillig* then *toves* are *slithy*.
2. Toves are *slithy*
3. Therefore, *toves* are *brillig*.

We know that the argument is invalid, even though we have no clue what "toves" are or what "slithy" or "brillig" means. The point is that we can identify formal fallacies without having to know what they mean.

Informal fallacies

Informal fallacies could be described as incorrect arguments where the error is not necessarily due to the form of the argument. Rather, it is the content or context that invalidates the arguments. Some informal fallacies include:

- Begging the question
- False dichotomy
- Equivocation
- Slippery slope fallacies

Begging the question

Consider the following argument:

Capital punishment is justified for crimes such as rape and murder because it is quite legitimate and appropriate for the state to put to death someone who has committed such heinous and inhuman acts.

The premise indicator "because" denotes the premise and (derivatively) the conclusion of this argument. In standard form, the argument is this:

1. It is legitimate and appropriate for the state to put to death someone who commits rape or murder.
2. Therefore, capital punishment is justified for crimes such as rape and murder.

You should notice something peculiar about this argument: the premise is essentially the same claim as the conclusion. The only difference is that the premise spells out what capital punishment means (the state putting criminals to death), whereas the conclusion just refers to capital punishment by name. Additionally, the premise

uses terms like "legitimate" and "appropriate," whereas the conclusion uses the related term, "justified." But these differences do not add up to any real differences in meaning. Thus, the premise is essentially saying the same thing as the conclusion. This is a problem: we want our premise to provide a reason for accepting the conclusion. But if the premise is the same claim as the conclusion, then it cannot possibly provide a reason for accepting the conclusion! Begging the question occurs when one (either explicitly or implicitly) assumes the truth of the conclusion in one or more of the premises.

Begging the question is thus a kind of circular reasoning. One interesting feature of this fallacy is that nothing is formally wrong with arguments of this form. Here is what I mean:

Consider an argument that explicitly commits the fallacy of begging the question. For example:

1. Capital punishment is morally permissible.
2. Therefore, capital punishment is morally permissible.

Now, apply any method of assessing validity to this argument, and you will see that it is valid by any method. If we use the informal test (by trying to imagine that the premises are true while the conclusion is false), then the argument passes the test, since any time the premise is true, the conclusion will have to be true as well (since it is the exact same statement). Likewise, the argument is valid by any formal test of validity, such as truth tables. But while this argument is technically valid, it is still a really bad argument. Why? Because the point of giving an argument in the first place is to provide some reason for thinking the conclusion is true for those who do not already accept the conclusion. But if one does not already accept the conclusion, then simply restating the conclusion in a different way is not going to convince them. Rather, a good argument will provide some reason for accepting the conclusion that is sufficiently independent of that conclusion itself. Begging the question utterly fails to do this, and this is why it counts as an

informal fallacy. What is interesting about begging the question is that absolutely nothing is formally wrong with the argument.

Whether or not an argument begs the question is not always an easy matter to sort out. As with all informal fallacies, detecting it requires a careful understanding of the meaning of the statements involved in the argument. Here is an example of an argument where it is not as clear whether there is a begging the question fallacy:

Christian belief is warranted because according to Christianity, there exists a being called "the Holy Spirit" who reliably guides Christians towards the truth regarding the central claims of Christianity.

One might think that there is a kind of circularity (or begging the question) involved in this argument since the argument appears to assume the truth of Christianity in justifying the claim that Christianity is true. As this example illustrates, the issue of whether an argument begs the question requires us to draw on our general knowledge of the world. This is the mark of an informal, rather than formal, fallacy.

False dichotomy

Suppose I was to argue the following:

1. Raising taxes on the wealthy will either hurt the economy or help it.
2. But it will not help the economy. Therefore, it will hurt the economy.

The standard form of this argument is:

1. Raising taxes on the wealthy will either hurt the economy or help it.
2. Raising taxes on the wealthy will not help the economy.

3. Therefore, raising taxes on the wealthy will hurt the economy.

This argument contains a fallacy called a "false dichotomy." A false dichotomy is simply a disjunction that does not exhaust all of the possible options. In this case, the problematic disjunction is the first premise: raising the taxes on the wealthy will either hurt the economy or help it. But these are not the only options. Another option is that raising taxes on the wealthy will have no effect on the economy.

You Are Either with Us or Against Us!

Figure 4.2 *George W. Bush addressed the General Assembly of the United Nations on 12 September 2002* Jessica/Feb 20: *Image license and source missing*

In a speech made on April 5, 2004, President Bush made the following remarks about the causes of the Iraq war:

Saddam Hussein once again defied the demands of the world. And so, I had a choice: Do I take the word of a madman, do I trust a person who had used weapons of mass destruction on his own people, plus people in the neighbourhood, or do I take the steps necessary to defend the country? Given that choice, I will defend America every time. (Peters & Woolley, n.d.)

The false dichotomy here is the claim that: either I trust the word of a madman, or I defend America (by going to war against Saddam Hussein's regime). The problem is that these are not the only options. Other options include ongoing diplomacy and economic sanctions. Thus, even if it true that Bush should not have trusted the word of Hussein, it does not follow that the only other option is going to war against Hussein's regime. (Furthermore, it is not clear how this was needed to defend America.) That is a false dichotomy. As with all the previous informal fallacies we have considered, the false dichotomy fallacy requires an understanding of the concepts involved. Thus, we have to use our understanding of world in order to assess whether a false dichotomy fallacy is being committed or not.

The arguments and efforts made by advocates for teaching creationism or intelligent design in science classrooms also adhere to this kind of thinking. As is evident in the United States, the goal of Christian Evangelicals in America is to infuse their form of Christianity into public schools. However, their tactics often involve trying to "debunk" Darwin's theory of evolution by natural selection. This comes from the false dichotomy that if Darwinism is false, then the Biblical creation myth must be true. It does not occur to them

that there are thousands upon thousands of creation myths around the world. Falsifying Darwin's theory does not automatically lead to people turning to the creation story in the bible.

Equivocation

Consider the following argument:

1. Children are a headache. Aspirin will make headaches go away.
2. Therefore, aspirin will make children go away.

This is a silly argument, but it illustrates the fallacy of equivocation. The problem is that the word "headache" is used equivocally—that is, in two different senses. In the first premise, "headache" is used figuratively, whereas in the second premise, "headache" is used literally. The argument is only successful if the meaning of "headache" is the same in both premises. But it is not here, so this argument is an instance of the equivocation fallacy.

Here is another example:

1. Taking a logic class helps you learn how to argue. But there is already too much hostility in the world today, and the fewer arguments, the better.
2. Therefore, you should not take a logic class.

In this example, the words "argue" and "argument" are used equivocally. The fallacy of equivocation is not always so easy to spot.

Here is a trickier example:

1. The existence of laws depends on the existence of intelligent beings, like humans, who create the laws.
2. However, some laws existed before there were any humans (e.g., laws of physics).
3. Therefore, there must be some non-human, intelligent being

that created these laws of nature.

The term "law" is used equivocally here. In the first premise, it is used to refer to societal laws, such as criminal law; in the second premise, it is used to refer to the laws of nature. Although we use the term "law" to apply to both cases, they are undeniably different. Societal laws, such as the criminal law of a society, are enforced by people, and there are punishments for breaking these laws. Natural laws, such as the laws of physics, cannot be broken, and thus there are no punishments for breaking them. (Does it make sense to scold the electron for not doing what the law says it will do?)

As with every informal fallacy we have examined in this section, equivocation can only be identified by understanding the meanings of the words involved. In fact, the definition of the equivocation fallacy refers to this very fact: the same word is being used in two different senses (i.e., with two different meanings). So, unlike formal fallacies, identifying the equivocation fallacy requires us to draw on our understanding of the meaning of words and of the world in general.

Slippery Slope Fallacies

Slippery slope fallacies depend on the concept of vagueness. When a concept or claim is vague, it means that we do not know precisely what claim is being made, or what the boundaries of the concept are. The classic example used to illustrate vagueness is the "sorites paradox." The term "sorites" is the Greek term for "heap" and the paradox comes from ancient Greek philosophy. Here is the paradox. I will give you two claims that each sound very plausible, but in fact lead to a paradox.

Here are the two claims:

1. One grain of sand is not a heap of sand.

2. If I start with something that is not a heap of sand, then adding one grain of sand to that will not create a heap of sand.

For example, two grains of sand is not a heap, thus (by the second claim) neither is three grains of sand. But since three grains of sand is not a heap then (by the second claim again) neither is four grains of sand. You can probably see where this is going. By continuing to add one grain of sand over and over, I will eventually end up with something that is clearly a heap of sand, but that will not count as a heap of sand if we accept both claims 1 and 2 above.

Philosophers continue to argue and debate about how to resolve the sorites paradox, but the point for us is just to illustrate the concept of vagueness. The concept "heap" is a vague concept in this example. But so are so many other concepts, such a colour concepts (red, yellow, green, etc.), moral concepts (right, wrong, good, bad), and just about any other concept you can think of. The one domain that seems unaffected by vagueness is mathematical and logical concepts. There are two fallacies related to vagueness: the causal slippery slope and the conceptual slippery slope. We will cover the conceptual slippery slope first since it relates most closely to the concept of vagueness I have explained above.

Conceptual Slippery Slope

It may be true that there is no essential difference between 499 grains of sand and 500 grains of sand. But even if that is so, it does not follow that there is no difference between 1 grain of sand and 5 billion grains of sand. In general, just because we cannot draw a distinction between A and B and between B and C, it does not mean we cannot draw a distinction between A and C. Here is an example of a conceptual slippery slope fallacy:

It is illegal for anyone under 21 to drink alcohol. But there is no difference between someone who is 21 and someone who is 20 years

11 months old. So, there is nothing wrong with someone who is 20 years and 11 months old drinking. But since there is no real distinction between being one month older and one month younger, there should not be anything wrong with drinking at any age. Therefore, there is nothing wrong with allowing a 10-year-old to drink alcohol.

Imagine the life of an individual in stages of 1-month intervals. Even if it is true that there is no distinction in kind between any one of those stages, it does not follow that there is no distinction between the extremes of either end. Clearly, there is a difference between a 5-year-old and a 25-year-old—a distinction, in kind, that is relevant to whether they should be allowed to drink alcohol. The conceptual slippery slope fallacy assumes that because we cannot draw a distinction between adjacent stages, we cannot draw a distinction at all between any stages.

One clear way of illustrating this is with colour. Think of a colour spectrum from purple to red to orange to yellow to green to blue. Each colour grades into the next without any distinguishable boundaries between the colours—a continuous spectrum. Even if it is true that for any two adjacent hues on the colour wheel, we cannot distinguish between the two, it does not follow from this that there is no distinction to be drawn between any two portions of the colour wheel, because then we would be committed to saying that there is no distinguishable difference between purple and yellow! The colour spectrum example illustrates the general point that just because the boundaries between very similar things on a spectrum are vague, it does not follow that there are no differences between any two things on that spectrum.

Whether or not one will identify an argument as committing a conceptual slippery slope fallacy depends on the other things one believes about the world. Thus, whether or not a conceptual slippery slope fallacy has been committed will often be a matter of some debate. It will itself be vague. Here is a good example that illustrates this point:

1. People are found not guilty by reason of insanity when they

cannot avoid breaking the law.
2. But people who are brought up in certain deprived social circumstances are not much more able than the legally insane to avoid breaking the law.
3. So, we should not find such individuals guilty any more than those who are legally insane.

Whether there is conceptual slippery slope fallacy here depends on what you think about a host of other things, including individual responsibility, free will, the psychological and social effects of deprived social circumstances, such as poverty, lack of opportunity, abuse, etc. Some people may think that there are big differences between those who are legally insane and those who grow up in deprived social circumstances. Others may not think the differences are so great. The issues here are subtle, sensitive, and complex, which is why it is difficult to determine whether there is any fallacy here or not. If the differences between those who are insane and those who are the product of deprived social circumstances turn out to be like the differences between one shade of yellow and an adjacent shade of yellow, then there is no fallacy here. But if the differences turn out to be analogous to those between yellow and green (i.e., with many distinguishable stages of difference in between), then there would indeed be a conceptual slippery slope fallacy here. The difficulty of distinguishing instances of the conceptual slippery slope fallacy, and the fact that distinguishing it requires us to draw on our knowledge about the world, shows that the conceptual slippery slope fallacy is an informal fallacy.

Causal Slippery Slope Fallacy

The causal slippery slope fallacy is committed when one event is said to lead to some other (usually disastrous) event via a chain of intermediary events. If you have ever seen Direct TV's "get rid of

cable" commercials, you will know exactly what I am talking about. (If you do not know what I am talking about, you should Google it right now and find out. They are quite funny.) Here is an example of a causal slippery slope fallacy (it is adapted from one of the Direct TV commercials):

If you use cable, your cable will probably go on the fritz. If your cable is on the fritz, you will probably get frustrated. When you get frustrated, you will probably hit the table. When you hit the table, your young daughter will probably imitate you. When your daughter imitates you, she will probably get thrown out of school. When she gets thrown out of school, she will probably meet undesirables. When she meets undesirables, she will probably marry undesirables. When she marries undesirables, you will probably have a grandson with a dog collar. Therefore, if you use cable, you will probably have a grandson with dog collar.

This example is silly and absurd, yes. But it illustrates the causal slippery slope fallacy. Slippery slope fallacies are always made up of a series of conjunctions of probabilistic conditional statements that link the first event to the last event. A causal slippery slope fallacy is committed when one assumes that just because each individual conditional statement is probable, the conditional that links the first event to the last event is also probable. Even if we grant that each "link" in the chain is individually probable, it does not follow that the whole chain (or the conditional that links the first event to the last event) is probable. For the sake of the argument, suppose we assign probabilities to each "link" or conditional statement, like in the following list. (I have italicized the consequents of the conditionals and assigned high conditional probabilities to them. The high probability is for the sake of the argument; I do not actually think these things are as probable as I have assumed here.)

- If you use cable, then *your cable will probably go on the fritz* (.9)
- If your cable is on the fritz, then *you will probably get angry* (.9)
- If you get angry, then *you will probably hit the table* (.9)
- If you hit the table, *your daughter will probably imitate you* (.8)

- If your daughter imitates you, *she will probably be kicked out of school* (.8)
- If she is kicked out of school, *she will probably meet undesirables* (.9)
- If she meets undesirables, *she will probably marry undesirables* (.8)
- If she marries undesirables, *you will probably have a grandson with a dog collar* (.8)

However, even if we grant the probabilities of each link in the chain are high (80–90% probable), the conclusion does not even reach a probability higher than chance. In order to figure the probability of a conjunction, we must multiply the probability of each conjunct:

(.9) × (.9) × (.9) × (.8) × (.8) × (.9) × (.8) × (.8) = .27

That means the probability of the conclusion (i.e., that if you use cable, you will have a grandson with a dog collar) is only 27%, despite the fact that each conditional has a relatively high probability! The causal slippery slope fallacy is actually a formal probabilistic fallacy. What makes it a formal, rather than informal, fallacy is that we can identify it without even having to know what the sentences of the argument mean. I could just have easily written out a nonsense argument comprised of a series of probabilistic conditional statements. But I would still have been able to identify the causal slippery slope fallacy because I would have seen that there was a series of probabilistic conditional statements leading to a claim that the conclusion of the series was also probable. That is enough to tell me that there is a causal slippery slope fallacy, even if I do not really understand the meanings of the conditional statements.

Media Attributions

Jessica/Feb 20: Information missing

Relevance Fallacies

What all relevance fallacies have in common is that they make an argument or response to an argument that is irrelevant. Relevance fallacies can be compelling psychologically, but it is important to distinguish between rhetorical techniques that are psychologically compelling, on the one hand, and rationally compelling arguments, on the other. What makes something a fallacy is that it fails to be rationally compelling once we have carefully considered it. That said, arguments that fail to be rationally compelling may still be psychologically or emotionally compelling. The first fallacy of relevance that we will consider, the ad hominem fallacy, is an excellent example a fallacy that can be psychologically compelling.

Ad Hominem

"Ad hominem" is a Latin phrase that can be translated into English as "against the man." In an ad hominem fallacy, instead of responding to (or attacking) the argument a person has made, one attacks the person themself. In short, one attacks the person making the argument rather than the argument itself. Here is an anecdote that reveals an ad hominem fallacy (which has actually occurred in my ethics class before):

A *philosopher named Peter Singer had made an argument that it is morally wrong to spend money on luxuries for oneself rather than give all of your money that you do not strictly need away to charity. The argument is actually an argument from analogy, but the essence of the argument is that, in this world, children who die preventable deaths every day, and there are charities who could save the lives of these children if they are funded by individuals from wealthy countries like our own. Since there are things that we all regularly buy that we do*

not need (e.g., Starbuck's lattes, beer, movie tickets, or extra clothes or shoes we do not really need), if we continue to purchase those things rather than using that money to save the lives of children, then we are essentially contributing to the deaths of those children.

In response to Singer's argument, one student in the class asked: "Does Peter Singer give his money to charity? Does he do what he says we are all morally required to do?" The implication of this student's question (which I confirmed by following up with her) was that if Peter Singer himself does not donate all his extra money to charities, then his argument is not any good and can be dismissed. But that would be to commit an ad hominem fallacy. Instead of responding to the argument that Singer had made, this student attacked Singer himself. That is, they wanted to know how Singer lived and whether he was a hypocrite or not. Was he the kind of person who would tell us all that we had to live a certain way but fail to live that way himself?

But all of this is irrelevant to assessing Singer's argument. Suppose that Singer did not donate his excess money to charity and instead spent it on luxurious things for himself. Still, the argument that Singer has given can be assessed on its own merits. Even if it were true that Peter Singer was a total hypocrite, his argument may nevertheless be rationally compelling. And it is the quality of the argument that we are interested in, not Peter Singer's personal life and whether or not he is hypocritical. Whether Singer is or is not a hypocrite is irrelevant to whether the argument he has put forward is strong or weak, valid or invalid. The argument stands on its own, and it is that argument, rather than Peter Singer himself, that we need to assess.

Nonetheless, there is something psychologically compelling about the question: Does Peter Singer practice what he preaches? I think what makes this question seem compelling is that humans are very interested in finding "cheaters" or hypocrites—those who say one thing and then do another. Evolutionarily, our concern with cheaters makes sense because cheaters cannot be trusted, and it is essential for us (as a group) to be able to pick out these people.

That said, whether or not a person giving an argument is a hypocrite is irrelevant to whether that person's argument is good or bad. So, there may be psychological reasons why humans are prone to find certain kinds of ad hominem fallacies psychologically compelling, even though ad hominem fallacies are not rationally compelling.

Not every instance in which someone attacks a person's character is an ad hominem fallacy. Suppose a witness is on the stand testifying against a defendant in a court of law. When the witness is cross examined by the defense lawyer, the defense lawyer tries to go for the witness's credibility, perhaps by digging up things about the witness's past. For example, the defense lawyer may find out that the witness cheated on her taxes five years ago or that the witness failed to pay her parking tickets. The reason this is not an ad hominem fallacy is that, in this case, the lawyer is trying to establish whether what the witness is saying is true or false and in order to determine that we have to know whether the witness is trustworthy. These facts about the witness's past may be relevant to determining whether we can trust the witness's word. In this case, the witness is making claims that are either true or false rather than giving an argument.

In contrast, when we are assessing someone's argument, the argument stands on its own in a way the witness's testimony does not. In assessing an argument, we want to know whether the argument is strong or weak, and we can evaluate the argument using the logical techniques surveyed in this text. In contrast, when a witness is giving a testimony, they are not trying to argue anything. Rather, they are simply making a claim about what did or did not happen. So, although it may seem like a lawyer is committing an ad hominem fallacy in bringing up things about the witness's past, these things are actually relevant to establishing the witness's credibility. In contrast, when considering an argument that has been given, we do not have to establish the arguer's credibility because we can assess their argument on its own merits. The arguer's personal life is irrelevant.

Straw Man

Suppose that my opponent has argued for a position (position A), and in response to his argument, I give a rationally compelling argument against position B, which is related to position A, but is much less plausible (and thus much easier to refute). What I have just done is attacked a straw man—a position that "looks like" the target position but is actually not that position. When one attacks a straw man, one commits the straw man fallacy. The straw man fallacy misrepresents one's opponent's argument and is thus a kind of irrelevance. Here is an example:

Two candidates for political office in Colorado, Tom and Fred, are having an exchange in a debate in which Tom has laid out his plan for putting more money into healthcare and education and Fred has laid out his plan which includes earmarking more state money for building more prisons to create more jobs and, thus, strengthen Colorado's economy. Fred responds to Tom's argument that we need to increase funding to healthcare and education as follows: "I am surprised, Tom, that you are willing to put our state's economic future at risk by sinking money into these programs that do not help create jobs. You see, folks, Tom's plan will risk sending our economy into a tailspin, risking harm to thousands of Coloradans. On the other hand, my plan supports a healthy and strong Colorado and would never bet our state's economic security on idealistic notions that simply do not work when the rubber meets the road."

Fred has committed the straw man fallacy. Just because Tom wants to increase funding to healthcare and education does not mean he does not want to help the economy. Furthermore, increasing funding to healthcare and education does not mean that fewer jobs will be created. Fred has attacked a position that is not the position that Tom holds, but is, in fact, a much less plausible, easier to refute position. However, it would be silly for any political candidate to run on a platform that included "harming the economy." Presumably no political candidate would run on such a

platform. Nonetheless, this exact kind of straw man is ubiquitous in political discourse in our country.

As with other relevance fallacies, straw man fallacies can be compelling on some level, even though they are irrelevant. It may be that part of the reason we are taken in by straw man fallacies is that humans are prone to "demonize" the "other"—including those who hold a moral or political position different from our own. It is easy to think bad things about those with whom we do not regularly interact. And it is easy to forget that people who are different than us are still people just like us in all the important respects.

Many years ago, atheists were commonly thought of as highly immoral people, and stories about the horrible things that atheists did in secret circulated widely. People believed that these strange "others" were capable of the most horrible savagery. After all, they may have reasoned, if you do not believe there is a God holding us accountable, why be moral? The Jewish philosopher, Baruch Spinoza, was an atheist who lived in the Netherlands in the 17th century. He was accused of all sorts of things that were commonly believed about atheists. But he was, in fact, as upstanding and moral as any person you could imagine. The people who knew Spinoza knew better, but how could so many people be so wrong about Spinoza? I suspect that part of the reason is that since were very few atheists at the time (or at least very few people who actually admitted to it without being killed by the Church), very few people ever knowingly encountered an atheist. Because of this, the stories about atheists could proliferate without being challenged by the facts. I suspect the same kind of phenomenon explains why certain kinds of straw man fallacies proliferate. If you are a conservative and mostly only interact with other conservatives, you might be prone to holding lots of false beliefs about liberals. And so maybe you are less prone to notice straw man fallacies targeted at liberals because the false beliefs you hold about them incline you to see the straw man fallacies as true.

Tu Quoque

"Tu quoque" is a Latin phrase that can be translated into English as "you too" or "you, also." The tu quoque fallacy is a way of avoiding answering a criticism by bringing up a criticism of your opponent instead. For example, suppose that two political candidates, A and B, are discussing their policies and A brings up a criticism of B's policy. In response, B brings up her own criticism of A's policy rather than responding to A's criticism of her policy. Here, B has committed the tu quoque fallacy. The fallacy is best understood as a way to avoid answering a tough criticism that one may not have a good answer to. This kind of thing happens all the time in political discourse.

Tu quoque, as I have presented it, is fallacious when the criticism one raises is simply to avoid having to answer a difficult objection to one's argument or view. However, there are circumstances in which a tu quoque kind of response is not fallacious. If the criticism that A brings toward B is a criticism that equally applies not only to A's position but to any position, then B is right to point this fact out. For example, suppose that A criticizes B for taking money from special interest groups. In this case, B would be totally right (and there would be no tu quoque fallacy committed) to respond that not only does A take money from special interest groups, but every political candidate running for office does. That is just a fact of life in American politics today. So, A really has no criticism at all to B since everyone does what B is doing, and it is unavoidable in many ways. Thus, B could (and should) respond with a "you too" rebuttal, and in this case, that rebuttal is not a tu quoque fallacy.

Genetic Fallacy

The genetic fallacy occurs when one argues (or, more commonly, implies) that the origin of something (e.g., a theory, idea, policy,

etc.) is a reason for rejecting (or accepting) it. For example, suppose that Jack is arguing that we should allow physician-assisted suicide, and Jill responds that that idea was first used in Nazi Germany. Jill has just committed a genetic fallacy because she is implying that because the idea is associated with Nazi Germany, there must be something wrong with the idea itself. What she should have done instead is explain what, exactly, is wrong with the idea rather than simply assuming that there must be something wrong with it since it has a negative origin. The origin of an idea has nothing inherently to do with its truth or plausibility. Suppose that Hitler constructed a mathematical proof in his early adulthood (he did not but just suppose). The validity of that mathematical proof stands on its own; the fact that Hitler was a horrible person has nothing to do with whether the proof is good. Likewise, with any other idea: ideas must be assessed on their own merits and the origin of an idea is neither a merit nor demerit of the idea. Although genetic fallacies are most often committed when one associates an idea with a negative origin, it can also go the other way: one can imply that because the idea has a positive origin, the idea must be true or more plausible.

Suppose that Jill argues that the Golden Rule is a good way to live one's life because the Golden Rule originated with Jesus in the Sermon on the Mount (it did not, actually; even though Jesus does state a version of the Golden Rule). Jill has committed the genetic fallacy in assuming that the (presumed) fact that Jesus is the origin of the Golden Rule has anything to do with whether the Golden Rule is a good idea.

I will end with an example from William James's (1902) seminal work, *The Varieties of Religious Experience*. In that book (originally a set of lectures), James considers the idea that if religious experiences could be explained in terms of neurological causes, then the legitimacy of the religious experience is undermined. James, being a materialist who thinks that all mental states are physical states—ultimately, a matter of complex brain chemistry—says that the fact that any religious experience has a physical cause does not undermine the veracity of that experience.

Although he does not use the term explicitly, James states that claiming that the physical origin of some experience undermines the veracity of that experience is a genetic fallacy. According to James, origin is irrelevant for assessing the veracity of an experience. In fact, he thinks that religious dogmatists who take the origin of the Bible to be the word of God are making exactly the same mistake as those who think that a physical explanation of a religious experience would undermine its veracity. We must assess ideas for their merits, not their origins.

Appeal to Consequences

The appeal to consequences fallacy is like the reverse of the genetic fallacy: the genetic fallacy is the mistake of trying to assess the truth or reasonableness of an idea based on the origin of the idea, whereas the appeal to consequences fallacy is the mistake of trying to assess the truth or reasonableness of an idea based on the (typically negative) consequences of accepting that idea. For example, suppose that the results of a study revealed that there are IQ differences between different races (this is a fictitious example; there is no credible study that I know of). In debating the results of this study, one researcher claims that if we were to accept these results, it would lead to increased racism in our society, which is not tolerable. Therefore, these results must not be right since if they were accepted, it would lead to increased racism. The researcher who responded in this way has committed the appeal to consequences fallacy. Again, we must assess the study on its own merits. If there is something wrong with the study—some flaw in its design, for example—then that would be a relevant criticism of the study. However, the fact that the results of the study, if widely circulated, would have a negative effect on society is not a reason for rejecting these results as false. The consequences of an idea

(good or bad) are irrelevant to the truth or reasonableness of that idea.

Notice that the researchers, being convinced of the negative consequences of the study on society, might rationally choose not to publish the study (for fear of the negative consequences). This is totally fine and is not a fallacy. The fallacy is not in choosing not to publish something that could have adverse consequences but in claiming that the results themselves are undermined by the negative consequences they could have. The fact is, sometimes truth can have negative consequences and falsehoods can have positive consequences. This just goes to show that the consequences of an idea are irrelevant to the truth or reasonableness of an idea.

Appeal to Authority

In a society like ours, we have to rely on authorities to get on in life. For example, the things I believe about electrons are not things that I have ever verified for myself. Rather, I have to rely on the testimony and authority of physicists to tell me what electrons are like. Likewise, when there is something wrong with my car, I have to rely on a mechanic (since I lack that expertise) to tell me what is wrong with it. Such is modern life. So, there is nothing wrong with needing to rely on authority figures in certain fields (people with the relevant expertise in that field)—it is inescapable.

The problem comes when we invoke someone whose expertise is not relevant to the issue for which we are invoking it. For example, suppose that a group of doctors sign a petition to prohibit abortions, claiming that abortions are morally wrong. If Bob cites that fact that these doctors are against abortion, therefore abortion must be morally wrong, then Bob has committed the appeal to authority fallacy. The problem is that doctors are not authorities on what is morally right or wrong. Even if they are authorities on how the body works and how to perform certain procedures (such

as abortion), it does not follow that they are authorities on whether or not these procedures should be performed—the ethical status of these procedures.

It would be just as much an appeal to consequences fallacy if Melissa were to argue that since some other group of doctors supported abortion, that shows that it must be morally acceptable. In either case, since doctors are not authorities on moral issues, their opinions on a moral issue like abortion are irrelevant. In general, an appeal to authority fallacy occurs when someone takes what an individual says as evidence for some claim, when that individual has no particular expertise in the relevant domain (even if they do have expertise in some other, unrelated, domain).

Summary

In this chapter, we explored some of the basic criteria for critical thinking. We looked at eight steps in processing an idea and then differentiated between formal and informal fallacies. We saw that formal fallacies depend on the form of the argument, while informal fallacies often appeal to psychological states in the absence of context.

Key Takeaways

- Rational, objective thinking is a hallmark of science.
- The eight processes used in critical thinking are 1) ask questions and be willing to wonder, 2) define the problem, 3) examine the evidence, 4) analyze assumptions and biases, 5) avoid emotional reasoning, 6) avoid oversimplification, 7) consider other interpretations, and 8) tolerate uncertainty.
- A fallacy is a mistake in reasoning.
- A formal fallacy is an argument whose form is invalid.
- Informal fallacies are those that cannot be identified without understanding the concepts involved in the argument

Acknowledgements

Parts of this chapter were adapted from the following Open Education Resources:

- "1.1 Psychology as a Science" in *Psychology – 1st Canadian Edition* by Walters (2020), via Thompson Rivers University, is used under a CC BY-NC-SA 4.0 license.
- *Introduction to Logic and Critical Thinking* by Matthew J. Van Cleave (2016), via B.C. Open Collection, is used under a CC BY-NC-SA 4.0 license.

References

Dazzi, C., & Pedrabissi, L. (2009). Graphology and personality: An empirical study on validity of handwriting analysis. *Psychological Reports*, 105(3, Suppl.), 1255–1268. https://doi.org/10.2466/pr0.105.f.1255-1268

Downey, J. E. (1919). *Graphology and the psychology of handwriting*. Warwick & York.

James, W. (1902). *The varieties of religious experience: A study in human nature, being the Gifford Lectures on natural religion delivered at Edinburgh in 1901–1902*. Longman.

LaBianca, J., & Gibson, B. (n.d.). Here's what your handwriting says about you. *Reader's Digest*. Retrieved January 2, 2020, from https://web.archive.org/web/20200102062135/https://www.rd.com/advice/work-career/handwriting-analysis/

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2010). *50 Great myths of popular psychology: Shattering widespread misconceptions about human behavior*. Wiley-Blackwell.

Münsterberg, H. (1915). *Business psychology*. La Salle Extension University.

Peters, G., & Woolley, J. T. (n.d.). *George W. Bush, remarks on job training and the national economy in Charlotte, North Carolina*. The American Presidency Project. Retrieved September 26, 2025, from https://www.presidency.ucsb.edu/node/214686

Wade, C. (1995). Using writing to develop and assess critical thinking. *Teaching of Psychology*, 22(1), 24–28. https://psycnet.apa.org/doi/10.1207/s15328023top2201_8

V. THE SCIENTIFIC METHOD

Learning Objectives

- Review a general model of scientific research.
- Describe the stages of the scientific method.
- Explain the difference between independent and dependent variables
- Differentiate between experimental and non-experimental research

Introduction

"Therefore, the seeker after the truth is not one who studies the writings of the ancients and, following his natural disposition, puts his trust in them, but rather the one who suspects his faith in them and questions what he gathers from them, the one who submits to argument and demonstration, and not to the sayings of a human being whose nature is fraught with all kinds of imperfection and deficiency. The duty of the man who investigates the writings of scientists, if learning the truth is his goal, is to make himself an enemy of all that he reads, and … attack it from every side. He should also suspect himself as he performs his critical examination of it, so that he may avoid falling into either prejudice or leniency."

أبو علي، الحسن بن الحسن بن الهيثم —

Abū ʿAlī al-Ḥasan ibn al-Ḥasan ibn al-Haytham

Kitāb al-Manāẓir or Book of Optics

(Sabra, 1989)

The above quote by Alhasan (the Latin form of his name) is poignant for this chapter as he is credited with the first systematic use of the scientific method. He based his theories on vision and light on experimentation (*i'tibar*, اختبار) and mathematics. In this chapter, we will explore the basics of the scientific method and see how it can be used to explore nature.

The Scientific Method

An Abstract from the Journal: Psychological Science

> Taking notes on laptops rather than in longhand is increasingly common. Many researchers have suggested that laptop note taking is less effective than longhand note taking for learning. Prior studies have primarily focused on students' capacity for multitasking and distraction when using laptops. The present research suggests that even when laptops are used solely to take notes, they may still be impairing learning because their use results in shallower processing. In three studies, we found that students who took notes on laptops performed worse on conceptual questions than students who took notes longhand. We show that whereas taking more notes can be beneficial, laptop note takers' tendency to transcribe lectures verbatim rather than processing information and reframing it in their own words is detrimental to learning. (Mueler & Oppenheimer, 2014, p. 1159)

In this abstract, the researcher has identified a research question—about the effect of taking notes on a laptop on learning—and identified why it is worthy of investigation—because the practice is ubiquitous and may be harmful to learning.

In this chapter, we will have a broad overview of the various stages of the research process. These include finding a topic of investigation, reviewing the literature, refining your research

question and generating a hypothesis, designing and conducting a study, analyzing the data, coming to conclusions, and reporting the results.

A Model of Scientific Research

Figure 5.1 presents a simple model of scientific research. The researchers formulate a research question, conduct an empirical study designed to answer the question, analyze the resulting data, draw conclusions about the answer to the question, and publish the results so that they become part of the research literature (i.e., all the published research in that field). Because the research literature is one of the primary sources of new research questions, this process can be thought of as a cycle. New research leads to new questions, which lead to new research, and so on. Figure 3.1 also indicates that research questions can originate outside of this cycle either with informal observations or with practical problems that need to be solved. But even in these cases, the researcher would start by checking the research literature to see if the question had already been answered and refining it based on what previous research had already found.

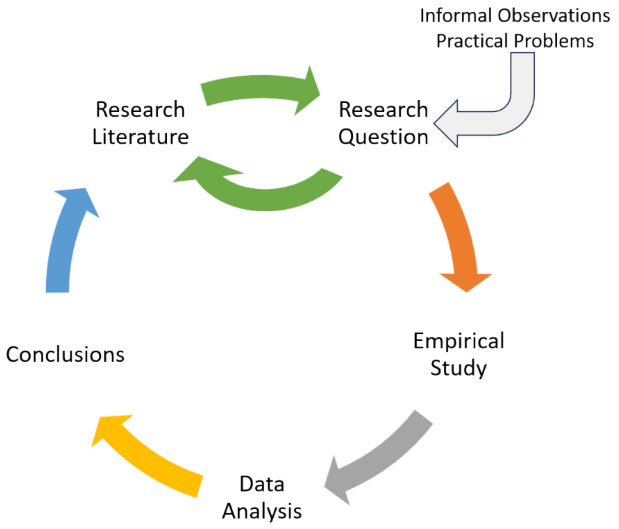


Figure 5.1 *A Simple Model of Scientific Research (Jhangiani et al., 2019)*
Jessica/Feb 20: Images license missing

A study by Mehl et al. (2007) is described nicely by this model. Their research question—whether women are more talkative than men—was suggested to them both by people's stereotypes and by claims published in the research literature about the relative talkativeness of women and men. When they checked the research literature, however, they found that this question had not been adequately addressed in scientific studies. They then conducted a careful empirical study, analyzed the results (finding very little difference between women and men), formed their conclusions, and published their work so that it became part of the research literature. The publication of their article is not the end of the story, however, because their work suggests many new questions (about the reliability of the result, about potential cultural differences, etc.) that will likely be taken up by them and by other researchers inspired by their work.

As another example, consider that as cell phones became more widespread during the 1990s, people began to wonder whether, and to what extent, cell phone use had a negative effect on driving. Many psychologists decided to tackle this question scientifically (e.g., Collet et al., 2010). It was clear from previously published research that engaging in a simple verbal task impairs performance on a perceptual or motor task carried out at the same time, but no one had studied the effect specifically of cell phone use on driving. Under carefully controlled conditions, these researchers compared people's driving performance while using a cell phone with their performance while not using a cell phone, both in the lab and on the road. They found that people's ability to detect road hazards, reaction time, and maintain control of the vehicle were all impaired by cell phone use. Each new study was published and became part of the growing research literature on this topic. For instance, other research teams subsequently demonstrated that cell phone conversations carry a greater risk than conversations with a passenger who is aware of driving conditions, which often become a point of conversation (e.g., Drews et al., 2004).

Theories and Hypotheses

Before describing how to develop a hypothesis, it is important to distinguish between a theory and a hypothesis. A theory is a coherent explanation or interpretation of one or more phenomena. Although theories can take a variety of forms, one thing they have in common is that they go beyond the phenomena they explain by including variables, structures, processes, functions, or organizing principles that have not been observed directly. Consider, for example, Zajonc's (1965) theory of social facilitation and social inhibition. He proposed that being watched by others while performing a task creates a general state of physiological arousal, which increases the likelihood of the dominant (most likely)

response. So, for highly practiced tasks, being watched increases the tendency to make correct responses, but for relatively unpractised tasks, being watched increases the tendency to make incorrect responses. Notice that this theory—which has come to be called drive theory—provides an explanation of both social facilitation and social inhibition that goes beyond the phenomena themselves by including concepts such as "arousal" and "dominant response," along with processes such as the effect of arousal on the dominant response.

Outside of science, referring to an idea as a theory often implies that it is untested—perhaps no more than a wild guess. In science, however, the term "theory" has no such implication. A theory is simply an explanation or interpretation of a set of phenomena. It can be untested, but it can also be extensively tested, well-supported, and accepted as an accurate description of the world by the scientific community. The theory of evolution by natural selection, for example, is a theory because it is an explanation of the diversity of life on earth—not because it is untested or unsupported by scientific research. On the contrary, the evidence for this theory is overwhelmingly positive and nearly all scientists accept its basic assumptions as accurate. Similarly, the "germ theory" of disease is a theory because it is an explanation of the origin of various diseases, not because there is any doubt that many diseases are caused by microorganisms that infect the body.

A hypothesis, on the other hand, is a specific prediction about a new phenomenon that should be observed if a particular theory is accurate. It is an explanation that relies on just a few key concepts. Hypotheses are often specific predictions about what will happen in a particular study. They are developed by considering existing evidence and using reasoning to infer what will happen in the specific context of interest. Hypotheses are often but not always derived from theories. So, a hypothesis is often a prediction based on a theory, but some hypotheses are a-theoretical; only after a set of observations have been made is a theory developed. This is because theories are broad in nature and explain larger bodies of

data. So, if our research question is really original, then we may need to collect some data and make some observations before we can develop a broader theory.

Theories and hypotheses always have this if-then relationship. "If drive theory is correct, then cockroaches should run through a straight runway faster, and a branching runway more slowly, when other cockroaches are present." Although hypotheses are usually expressed as statements, they can always be rephrased as questions. "Do cockroaches run through a straight runway faster when other cockroaches are present?" Thus, deriving hypotheses from theories is an excellent way of generating interesting research questions.

But how do researchers derive hypotheses from theories? One way is to generate a research question using the techniques discussed in this chapter and then ask whether any theory implies an answer to that question. For example, you might wonder whether expressive writing about positive experiences improves health as much as expressive writing about traumatic experiences. Although this question is an interesting one on its own, you might then ask whether the habituation theory—the idea that expressive writing causes people to habituate to negative thoughts and feelings—implies an answer. In this case, it seems clear that if the habituation theory is correct, then expressive writing about positive experiences should not be effective because it would not cause people to habituate to negative thoughts and feelings. A second way to derive hypotheses from theories is to focus on some component of the theory that has not yet been directly observed. For example, a researcher could focus on the process of habituation—perhaps hypothesizing that people should show fewer signs of emotional distress with each new writing session.

Among the very best hypotheses are those that distinguish between competing theories. For example, Norbert Schwarz and his colleagues considered two theories of how people make judgments about themselves, such as how assertive they are (Schwarz et al., 1991). Both theories held that such judgments are based on relevant examples that people bring to mind. However, one theory was that

people base their judgments on the number of examples they bring to mind, and the other was that people base their judgments on how easily they bring those examples to mind. To test these theories, the researchers asked people to recall either six times when they were assertive (which is easy for most people) or 12 times (which is difficult for most people). Then, they asked them to judge their own assertiveness. Note that the number-of-examples theory implies that people who recalled 12 examples should judge themselves to be more assertive because they recalled more examples, but the ease-of-examples theory implies that participants who recalled six examples should judge themselves as more assertive because recalling the examples was easier. Thus, the two theories made opposite predictions so that only one of the predictions could be confirmed. The surprising result was that participants who recalled fewer examples judged themselves to be more assertive—providing particularly convincing evidence in favor of the ease-of-retrieval theory over the number-of-examples theory.

Theory Testing

The primary way that scientific researchers use theories is sometimes called the hypothetico-deductive method (although this term is much more likely to be used by philosophers of science than by scientists themselves). Researchers begin with a set of phenomena and either construct a theory to explain or interpret them or choose an existing theory to work with. They then make a prediction about some new phenomenon that should be observed if the theory is correct. Again, this prediction is called a hypothesis. The researchers then conduct an empirical study to test the hypothesis. Finally, they re-evaluate the theory in light of the new results and revise it if necessary.

This process is usually conceptualized as a cycle because the researchers can then derive a new hypothesis from the revised

theory, conduct a new empirical study to test the hypothesis, and so on. As Figure 5.2 shows, this approach meshes nicely with the model of scientific research in psychology presented earlier in the textbook—creating a more detailed model of "theoretically motivated" or "theory-driven" research.

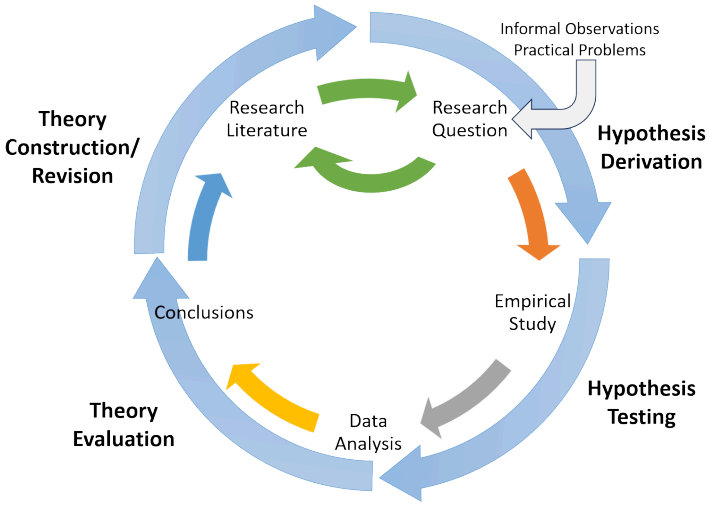


Figure 5.2 *Hypothetico-Deductive Method Combined with the General Model of Scientific Research in Psychology Together they form a model of theoretically motivated research.* *Jessica/Feb 20: Images license missing*

As an example, let us consider Zajonc's (1965) research on social facilitation and inhibition. He started with a somewhat contradictory pattern of results from the research literature. He then constructed his drive theory, according to which being watched by others while performing a task causes physiological arousal, which increases an organism's tendency to make the dominant response. This theory predicts social facilitation for well-learned tasks and social inhibition for poorly learned tasks. He now had a theory that organized previous results in a meaningful way—but he still needed to test it. He hypothesized that if his theory was correct, he should observe that the presence of others

improves performance in a simple laboratory task but inhibits performance in a difficult version of the very same laboratory task.

To test this hypothesis, one of the studies he conducted used cockroaches as subjects (Zajonc et al., 1969). The cockroaches ran either down a straight runway (an easy task for a cockroach) or through a cross-shaped maze (a difficult task for a cockroach) to escape into a dark chamber when a light was shined on them. They did this either while alone or in the presence of other cockroaches in clear plastic "audience boxes." Zajonc found that cockroaches in the straight runway reached their goal more quickly in the presence of other cockroaches, but cockroaches in the cross-shaped maze reached their goal more slowly when they were in the presence of other cockroaches. Thus, he confirmed his hypothesis and provided support for his drive theory. (Zajonc also showed that drive theory existed in humans [Zajonc & Sales, 1966] in many other studies afterward).

Incorporating Theory into Your Research

When you write your research report or plan your presentation, be aware that there are two basic ways that researchers usually include theories. The first is to raise a research question, answer that question by conducting a new study, and then offer one or more theories (usually more) to explain or interpret the results. This format works well for applied research questions and for research questions that existing theories do not address. The second way is to describe one or more existing theories, derive a hypothesis from one of those theories, test the hypothesis in a new study, and finally reevaluate the theory. This format works well when there is an existing theory that addresses the research question—especially if the resulting hypothesis is surprising or conflicts with a hypothesis derived from a different theory.

Using theories in your research will not only give you guidance

in coming up with experiment ideas and possible projects, but it will also lend legitimacy to your work. Psychologists have been interested in a variety of human behaviours and have developed many theories along the way. Using established theories will help you break new ground as a researcher, not limit you from developing your own ideas.

Characteristics of a Good Hypothesis

There are three general characteristics of a good hypothesis.

First, a good hypothesis must be testable and falsifiable. We must be able to test the hypothesis using the methods of science, and if you recall Popper's falsifiability criterion, it must be possible to gather evidence that will disconfirm the hypothesis if it is indeed false.

Second, a good hypothesis must be logical. As described above, hypotheses are more than just a random guess. Hypotheses should be informed by previous theories or observations and logical reasoning. Typically, we begin with a broad and general theory and use deductive reasoning to generate a more specific hypothesis to test based on that theory. Occasionally, however, when there is no theory to inform our hypothesis, we use inductive reasoning—which involves using specific observations or research findings—to form a more general hypothesis.

Finally, the hypothesis should be positive. That is, the hypothesis should make a positive statement about the existence of a relationship or effect, rather than a statement that a relationship or effect does not exist. As scientists, we do not set out to show that relationships do not exist or that effects do not occur, so our hypotheses should not be worded in a way to suggest that an effect or relationship does not exist. The nature of science is to assume that something does not exist and then seek to find evidence to prove this wrong to show that it really does exist. That may seem

backward to you but that is the nature of the scientific method. The underlying reason for this is beyond the scope of this chapter but it has to do with statistical theory.

Media Attributions

- **Figure 5.1** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University.
- **Figure 5.2** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University.

Designing a Research Study

Learning Objectives

- Define the concept of a variable, distinguish quantitative from categorical variables, and give examples of variables that might be of interest to psychologists.
- Explain the difference between a population and a sample.
- Distinguish between experimental and non-experimental research.
- Distinguish between lab studies, field studies, and field experiments.

Identifying and Defining the Variables and Population

Variables and Operational Definitions

Part of generating a hypothesis involves identifying the variables that you want to study and operationally defining those variables so that they can be measured. Research questions in psychology are about variables. A variable is a quantity or quality that varies

across people or situations. For example, the height of the students enrolled in a university course is a variable because it varies from student to student. The chosen major of the students is also a variable as long as not everyone in the class has declared the same major. Almost everything in our world varies and as such thinking of examples of constants (things that don't vary) is far more difficult. A rare example of a constant is the speed of light. Variables can be either quantitative or categorical. A quantitative variable is a quantity, such as height, that is typically measured by assigning a number to each individual. Other examples of quantitative variables include people's level of talkativeness, how depressed they are, and the number of siblings they have. A categorical variable is a quality, such as chosen major, and is typically measured by assigning a category label to each individual (e.g., Psychology, English, Engineering, etc.). Other examples include people's nationality, their occupation, and whether they are receiving psychotherapy.

After the researcher generates their hypothesis and selects the variables they want to manipulate and measure, the researcher needs to find ways to actually measure the variables of interest. This requires an operational definition—a definition of the variable in terms of precisely how it is to be measured. Most variables that researchers are interested in studying cannot be directly observed or measured and this poses a problem because empiricism (observation) is at the heart of the scientific method. Operationally defining a variable involves taking an abstract construct like depression that cannot be directly observed and transforming it into something that can be directly observed and measured. Most variables can be operationally defined in many different ways. For example, depression can be operationally defined as people's scores on a paper-and-pencil depression scale such as the Beck Depression Inventory, the number of depressive symptoms they are experiencing, or whether they have been diagnosed with major depressive disorder. Researchers are wise to choose an operational definition that has been used extensively in the research literature.

Sampling and Measurement

In addition to identifying which variables to manipulate and measure, and operationally defining those variables, researchers need to identify the population of interest. Researchers in psychology are usually interested in drawing conclusions about some very large group of people. This is called the population. It could be all American teenagers, children with autism, professional athletes, or even just human beings—depending on the interests and goals of the researcher. But they usually study only a small subset or sample of the population. For example, a researcher might measure the talkativeness of a few hundred university students with the intention of drawing conclusions about the talkativeness of men and women in general. It is important, therefore, for researchers to use a representative sample—one that is similar to the population in important respects.

One method of obtaining a sample is simple random sampling, in which every member of the population has an equal chance of being selected for the sample. For example, a pollster could start with a list of all the registered voters in a city (the population), randomly select 100 of them from the list (the sample), and ask those 100 whom they intend to vote for. Unfortunately, random sampling is difficult or impossible in most psychological research because the populations are less clearly defined than the registered voters in a city. How could a researcher give all American teenagers or all children with autism an equal chance of being selected for a sample? The most common alternative to random sampling is convenience sampling, in which the sample consists of individuals who happen to be nearby and willing to participate (such as introductory psychology students). Of course, the obvious problem with convenience sampling is that the sample might not be representative of the population and therefore it may be less appropriate to generalize the results from the sample to that population.

Designing a Research Study

Identifying and Defining the Variables and Population

Variables and Operational Definitions

Part of generating a hypothesis involves identifying the variables that you want to study and operationally defining those variables so that they can be measured. Research questions in psychology are about variables. A variable is a quantity or quality that varies across people or situations. For example, the height of the students enrolled in a university course is a variable because it varies from student to student. The chosen major of the students is also a variable as long as not everyone in the class has declared the same major. Almost everything in our world varies, and as such, thinking of examples of constants (things that do not vary) is far more difficult. A rare example of a constant is the speed of light.

Variables can be either quantitative or categorical. A quantitative variable is a quantity, such as height, that is typically measured by assigning a number to each individual. Other examples of quantitative variables include people's level of talkativeness, how depressed they are, and the number of siblings they have. A categorical variable is a quality, such as chosen major, and is typically measured by assigning a category label to each individual (e.g., Psychology, English, Engineering). Other examples include people's nationality, their occupation, and whether they are receiving psychotherapy.

After the researcher generates their hypothesis and selects the variables they want to manipulate and measure, the researcher

needs to find ways to actually measure the variables of interest. This requires an operational definition—a definition of the variable in terms of precisely how it is to be measured. Most variables that researchers are interested in studying cannot be directly observed or measured; this poses a problem because empiricism (observation) is at the heart of the scientific method. Operationally defining a variable involves taking an abstract construct, like depression, that cannot be directly observed and transforming it into something that can be directly observed and measured. Most variables can be operationally defined in many different ways. For example, depression can be operationally defined as people's scores on a paper-and-pencil depression scale, such as the Beck Depression Inventory, the number of depressive symptoms they are experiencing, or whether they have been diagnosed with major depressive disorder. Researchers are wise to choose an operational definition that has been used extensively in the research literature.

Sampling and Measurement

In addition to identifying which variables to manipulate and measure and operationally defining those variables, researchers need to identify the population of interest. Researchers in psychology are usually interested in drawing conclusions about some very large group of people. This is called the population. It could be all American teenagers, children with autism, professional athletes, or even just human beings—depending on the interests and goals of the researcher. But they usually study only a small subset or sample of the population. For example, a researcher might measure the talkativeness of a few hundred university students with the intention of drawing conclusions about the talkativeness of men and women in general. Therefore, it is important for researchers to use a representative sample—one that is similar to the population in important respects.

One method of obtaining a sample is simple random sampling, in which every member of the population has an equal chance of being selected for the sample. For example, a pollster could start with a list of all the registered voters in a city (the population), randomly select 100 of them from the list (the sample), and ask those 100 whom they intend to vote for. Unfortunately, random sampling is difficult or impossible in most psychological research because the populations are less clearly defined than the registered voters in a city. How could a researcher give all American teenagers or all children with autism an equal chance of being selected for a sample? The most common alternative to random sampling is convenience sampling, in which the sample consists of individuals who happen to be nearby and willing to participate (such as introductory psychology students). Of course, the obvious problem with convenience sampling is that the sample might not be representative of the population, and therefore, it may be less appropriate to generalize the results from the sample to that population.

Experimental vs. Non-Experimental Research

The next step a researcher must take is to decide which type of approach they will use to collect the data. As you will learn in your research methods course, there are many different approaches to research that can be divided in many different ways. One of the most fundamental distinctions is between experimental and non-experimental research.

Experimental Research

Researchers who want to test hypotheses about causal relationships between variables (i.e., their goal is to explain) need to use an experimental method. This is because the experimental method is the only method that allows us to determine causal relationships. Using the experimental approach, researchers first manipulate one or more variables while attempting to control extraneous variables, and then, they measure how the manipulated variables affect participants' responses.

The terms "independent variable" and "dependent variable" are used in the context of experimental research. The independent variable is the variable the experimenter manipulates (the presumed cause) and the dependent variable is the variable the experimenter measures (the presumed effect).

Extraneous variables are any variable other than the dependent variable. Confounds are a specific type of extraneous variable that systematically varies along with the variables under investigation and, therefore, provides an alternative explanation for the results. When researchers design an experiment, they need to ensure that

they control for confounds. They need to ensure that extraneous variables do not become confounding variables because, in order to make a causal conclusion, they need to make sure alternative explanations for the results have been ruled out.

As an example, if we manipulate the lighting in the room and examine the effects of that manipulation on workers' productivity, then the lighting conditions (bright lights vs. dim lights) would be considered the independent variable and the workers' productivity would be considered the dependent variable. If the bright lights are noisy, then that noise would be a confound since the noise would be present whenever the lights are bright and the noise would be absent when the lights are dim. If noise is varying systematically with light, then we would not know if a difference in worker productivity across the two lighting conditions is due to noise or light. So, confounds are bad—they disrupt our ability to make causal conclusions about the nature of the relationship between variables. However, if there is noise in the room both when the lights are on and when the lights are off, then noise is merely an extraneous variable (a variable other than the independent or dependent variable) and we do not worry much about extraneous variables. This is because unless a variable varies systematically with the manipulated independent variable, it cannot be a competing explanation for the results.

Non-Experimental Research

Researchers who are simply interested in describing characteristics of phenomena, describing relationships between variables, and using those relationships to make predictions can use non-experimental research. Using the non-experimental approach, the researcher simply measures variables as they naturally occur, but they do not manipulate them. For instance, if you just measured the number of traffic fatalities in America last year that involved

the use of a cell phone but did not actually manipulate cell phone use, then this would be categorized as non-experimental research. Alternatively, if you stood at a busy intersection and recorded drivers' genders and whether or not they were using a cell phone when they passed through the intersection to see whether men or women are more likely to use a cell phone when driving, then this would be non-experimental research.

It is important to point out that non-experimental does not mean non-scientific. Non-experimental research is scientific in nature. It can be used to fulfil two of the three goals of science (to describe and to predict). However, unlike with experimental research, we cannot make causal conclusions using this method; we cannot say that one variable causes another variable using this method.

Laboratory vs. Field Research

The next major distinction between research methods is between laboratory and field studies. A laboratory study is a study conducted in the laboratory environment. In contrast, a field study is a study conducted in the real-world, in a natural environment.

Laboratory experiments typically have high internal validity. Internal validity refers to the degree to which we can confidently infer a causal relationship between variables. When we conduct an experimental study in a laboratory environment, we have very high internal validity because we manipulate one variable while controlling all other outside extraneous variables. When we manipulate an independent variable, observe an effect on a dependent variable, and control for everything else so that the only difference between our experimental groups or conditions is the one manipulated variable, then we can be quite confident that it is the independent variable that is causing the change in the dependent variable. In contrast, because field studies are conducted in the real-world, the experimenter typically has less control over

the environment and potential extraneous variables; this decreases internal validity, making it less appropriate to arrive at causal conclusions.

But there is typically a trade-off between internal and external validity. External validity simply refers to the degree to which we can generalize the findings to other circumstances or settings, like the real-world environment. When internal validity is high, external validity tends to be low; and when internal validity is low, external validity tends to be high. So, laboratory studies are typically low in external validity, while field studies are typically high in external validity. Since field studies are conducted in the real-world environment, it is far more appropriate to generalize the findings to that real-world environment than when the research is conducted in the more artificial sterile laboratory.

Finally, there are field studies that are non-experimental in nature because nothing is manipulated. But there are also field experiments where an independent variable is manipulated in a natural setting and extraneous variables are controlled. Depending on their overall quality and the level of control of extraneous variables, such field experiments can have high external and high internal validity.

Summary

In this chapter, we discussed some of the basics of the scientific method. Regardless of discipline, all scientists use some version of the scientific method to understand nature. We use the scientific method to avoid biases, increase objectivity, and reach a consensus on our theories. However, the scientific method is not some ultimate truth or law of nature. It is an ever-evolving paradigm that has served us well so far. Scientists continue to refine it and explore its advantages and limitations.

Key Takeaways

- The scientific method is a paradigm that allows researchers to understand the world, reduce biases, increase objectivity, and develop explanatory theories.
- Researchers formulate a research question, conduct an empirical study designed to answer the question, analyze the resulting data, draw conclusions about the answer to the question, and publish the results.
- A theory is a coherent explanation or interpretation of one or more phenomena.
- A hypothesis is a specific prediction about a new phenomenon that should be observed if a particular theory is accurate.
- Researchers who want to test hypotheses about causal relationships between variables need to use an

experimental method.

- The independent variable is the variable the experimenter manipulates (the presumed cause), and the dependent variable is the variable the experimenter measures (the presumed effect).
- Researchers who are simply interested in describing characteristics of phenomena, describing relationships between variables, and using those relationships to make predictions can use non-experimental research

Acknowledgements

Parts of this chapter were adapted from the following open education resources:

The chapter "12. Analyzing the Data" in *Research Methods in Psychology* by Jhangiani et al. (2019), via Kwantlen Polytechnic University is used under a CC BY-NC-SA 4.0 license.

The chapter "3. Scientific Method" in *Introduction to History and Philosophy of Science* by Barseghyan et al. (2018), via eCampusOntario, is used under a CC BY 4.0 license.

References

References

Barseghyan, H., Overgaard, N., & Rupik, G. (2018). *Introduction to history and philosophy of science*. eCampusOntario. https://ecampusontario.pressbooks.pub/introhps/

Collet, C., Guillot, A., & Petit, C. (2010). Phoning while driving I: A review of epidemiological, psychological, behavioural and physiological studies. *Ergonomics*, 53(5), 589–601. https://doi.org/10.1080/00140131003672023

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2004). Passenger and cell-phone conversations in simulated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(19), 2210–2212. https://doi.org/10.1177/154193120404801901

Jhangiani, R. S., Chiang, I-C. A., Cuttler, C., & Leighton, D. C. (2019). *Research methods in psychology* (4th ed.). Kwantlen Polytechnic University. https://kpu.pressbooks.pub/psychmethods4e/

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men?. *Science (New York, N.Y.)*, 317(5834), 82. https://doi.org/10.1126/science.1139940

Mueller, P.A., & Oppenheimer, D.M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*, 25(6), 1159-1168. https://doi.org/10.1177/0956797614524581

Sabra, A. I., Ed. & Trans. (1989). The optics of Ibn al-Haytham: Books I–III on direct vision (Vols. 1–2). The Warburg Institute.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social*

Psychology, 61(2), 195–202. https://doi.org/10.1037/0022-3514.61.2.195

Zajonc, R. B. (1965). Social facilitation: A solution is suggested for an old unresolved social psychological problem. *Science*, 149(3681), 269–274. https://doi.org/10.1126/science.149.3681.269

Zajonc, R. B., Heingartner, A., & Herman, E. M. (1969). Social enhancement and impairment of performance in the cockroach. *Journal of Personality and Social Psychology*, 13(2), 83–92. https://doi.org/10.1037/h0028063

Zajonc, R. B. & Sales, S. M. (1966). Social facilitation of dominant and subordinate responses. *Journal of Experimental Social Psychology*, 2(2), 160–168. https://doi.org/10.1016/0022-1031(66)90077-1

VI. DESCRIPTIVE RESEARCH

Learning Objectives

- List several ways in which qualitative research differs from quantitative research in psychology.
- Describe the strengths and weaknesses of qualitative research in psychology compared with quantitative research.
- Give examples of qualitative research in psychology.
- List the various types of observational research methods and distinguish between each.
- Describe the strengths and weakness of each observational research method.

Introduction

ἆρ᾽ οὖν οὐχ ὑπάρχειν δεῖ τοῖς εὖ γε καὶ καλῶς ῥηθησομένοις τὴν τοῦ λέγοντος διάνοιαν εἰδυῖαν τὸ ἀληθὲς ὧν ἂν ἐρεῖν πέρι μέλλῃ;
If a speech is to be good, must not the mind of the speaker know the truth about the matters of which he is to speak?

— Socrates, *Phaedo* (260a)

Science is not a single set of instructions on how to understand the world around us. It is a set of basic ideas implemented through numerous evolving methodologies. Part of thinking critically about science involves understanding the different ways scientists collect and analyse information. This chapter will introduce you to descriptive or qualitative research methods. We will begin by understanding the definition of qualitative research and then go on to explore different types of qualitative research.

Qualitative Research

This book talks a lot about quantitative research, in part because most scientific studies are quantitative in nature. Quantitative researchers typically start with a focused research question or hypothesis, collect a small amount of numerical data from a large number of individuals, describe the resulting data using statistical techniques, and draw general conclusions about some large population. Although this method is by far the most common approach to conducting empirical research, there is an important alternative called qualitative research.

Qualitative research originated in the disciplines of anthropology and sociology but is now used to study a variety of topics. Qualitative researchers generally begin with a less focused research question, collect large amounts of relatively "unfiltered" data from a relatively small number of individuals, and describe their data using nonstatistical techniques, such as grounded theory, thematic analysis, critical discourse analysis, or interpretative phenomenological analysis. They are usually less concerned with drawing general conclusions about human behaviour than with understanding the experience of their research participants in detail.

A *Qualitative Study*

Consider, for example, a study by researcher Per Lindqvist and his colleagues (2008), who wanted to learn how the families of teenage suicide victims cope with their loss. They did not have a specific research question or

hypothesis, such as "What percentage of family members join suicide support groups?" Instead, they wanted to understand the variety of reactions that families had, with a focus on what it is like from their perspectives.

To address this question, they interviewed the families of 10 teenage suicide victims in their homes in rural Sweden. The interviews were relatively unstructured, beginning with a general request for the families to talk about the victim and ending with an invitation to talk about anything else that they wanted to tell the interviewer. One of the most important themes that emerged from these interviews was that even as life returned to "normal," the families continued to struggle with the question of why their loved one committed suicide. This struggle appeared to be especially difficult for families in which the suicide was most unexpected.

The Purpose of Qualitative Research

The strength of quantitative research is its ability to provide precise answers to specific research questions and draw general conclusions about human behaviour. This method is how we know that people have a strong tendency to obey authority figures, for example, and that female undergraduate students are not substantially more talkative than male undergraduate students. But while quantitative research is good at providing precise answers to specific research questions, it is not nearly as good at generating novel and interesting research questions. Likewise, while quantitative research is good at drawing general conclusions about human behaviour, it is not nearly as good at providing detailed

descriptions of the behaviour of particular groups in particular situations. And quantitative research is not very good at communicating what it is actually like to be a member of a particular group in a particular situation.

But the relative weaknesses of quantitative research are the relative strengths of qualitative research. Qualitative research can help researchers generate new and interesting research questions and hypotheses. The research of Lindqvist and colleagues (2008), for example, suggests that there may be a general relationship between how unexpected a suicide is and how consumed the family is with trying to understand why the teen committed suicide. This relationship can now be explored using quantitative research. But it is unclear whether this question would have arisen at all without the researchers sitting down with the families and listening to what they themselves wanted to say about their experience.

Qualitative research can also provide rich and detailed descriptions of human behaviour in the real-world contexts in which it occurs. Among qualitative researchers, this depth is often referred to as "thick description" (Geertz, 1973). Similarly, qualitative research can convey a sense of what it is actually like to be a member of a particular group or in a particular situation—what qualitative researchers often refer to as the "lived experience" of the research participants. Lindqvist and colleagues (2008), for example, describe how all the families spontaneously offered to show the interviewer the victim's bedroom or the place where the suicide occurred—revealing the importance of these physical locations to the families. It seems unlikely that a quantitative study would have discovered this detail.

Table 6.1 Qualitative vs. Quantitative

Qualitative	Quantitative
In-depth information about relatively few people	Less depth information with larger samples
Conclusions based on interpretations drawn by the investigator	Conclusions based on statistical analyses
Global and exploratory	Specific and focused

Data Collection and Analysis in Qualitative Research

Data collection approaches in qualitative research are quite varied and can involve naturalistic observation, participant observation, archival data, artwork, and many other things. But one of the most common approaches, especially for psychological research, is to conduct interviews. Interviews in qualitative research can be unstructured—consisting of a small number of general questions or prompts that allow participants to talk about what is of interest to them—or structured—where there is a strict script that the interviewer does not deviate from. Most interviews are in between the two and are called semi-structured interviews, where the researcher has a few consistent questions and can follow up by asking more detailed questions about the topics that come up. Such interviews can be lengthy and detailed, but they are usually conducted with a relatively small sample. The unstructured interview was the approach used by Lindqvist and colleagues (2008) in their research on the families of suicide victims because the researchers were aware that how much was disclosed about such a sensitive topic should be led by the families, not by the researchers.

Another approach used in qualitative research involves small groups of people who participate together in interviews focused on a particular topic or issue, known as focus groups. The interaction

among participants in a focus group can sometimes bring out more information than can be learned in a one-on-one interview. The use of focus groups has become a standard technique in business and industry among those who want to understand consumer tastes and preferences. The content of all focus group interviews is usually recorded and transcribed to facilitate later analyses. However, we know from social psychology that group dynamics are often at play in any group, including focus groups, and it is useful to be aware of those possibilities. For example, the desire to be liked by others can lead participants to provide inaccurate answers that they believe will be perceived favorably by the other participants.

A Focus Group Example

Consider a focus group that is talking about work habits. Let us say that one of the individuals is highly extraverted and outgoing. What problems do you anticipate (if any) in having a balanced discussion in such a focus group?

Data Analysis in Qualitative Research

Although quantitative and qualitative research generally differ along several important dimensions (e.g., the specificity of the research question, the type of data collected), it is the method of data analysis that distinguishes them more clearly than anything else. To illustrate this idea, imagine a team of researchers that conducts a series of unstructured interviews with people recovering from alcohol use disorder to learn about the role of their religious faith

in their recovery. Although this project sounds like qualitative research, imagine that once they collect the data, they code the data in terms of how often each participant mentions God (or a "higher power") and then use descriptive and inferential statistics to find out whether those who mention God more often are more successful in abstaining from alcohol. Now, it sounds like quantitative research. In other words, the quantitative-qualitative distinction depends more on what researchers do with the data they have collected than with why or how they collected the data.

But what does qualitative data analysis look like? Just as there are many ways to collect data in qualitative research, there are many ways to analyze data. Here, we focus on one general approach called grounded theory (Glaser & Strauss, 1967). This approach was developed within the field of sociology in the 1960s and has gradually gained popularity in psychology. Remember that in quantitative research, it is typical for the researcher to start with a theory, derive a hypothesis from that theory, and then collect data to test that specific hypothesis.

In qualitative research using grounded theory, researchers start with the data and develop a theory or an interpretation that is "grounded in" those data. They do this analysis in stages. First, they identify ideas that repeat throughout the data. Then, they organize these ideas into a smaller number of broader themes. Finally, they write a theoretical narrative—an interpretation of the data in terms of the themes that they have identified. This theoretical narrative focuses on the subjective experience of the participants and is usually supported by many direct quotations from the participants themselves.

As an example, consider a study by researchers Laura Abrams and Laura Curran (2009), who used the grounded theory approach to study the experience of postpartum depression symptoms among low-income mothers. Their data were the result of unstructured interviews with 19 participants. Table 6.2 shows the five broad themes the researchers identified and the more specific repeating ideas that made up each of those themes. In their research report,

they provide numerous quotations from their participants, such as this one from "Destiny":

> Well, just recently my apartment was broken into and the fact that his Medicaid for some reason was cancelled so a lot of things was happening within the last two weeks all at one time. So that in itself I don't want to say almost drove me mad but it put me in a funk....Like I really was depressed. (p. 357)

Their theoretical narrative focused on the participants' experience of their symptoms, not as an abstract "affective disorder" but as closely tied to the daily struggle of raising children alone under often difficult circumstances (Table 6.2).

Table 6.2 Repeating Ideas in Themes

Theme	Repeating ideas
Ambivalence	"I wasn't prepared for this baby," "I didn't want to have any more children."
Caregiving overload	"Please stop crying," "I need a break," "I can't do this anymore."
Juggling	"No time to breathe," "Everyone depends on me," "Navigating the maze."
Mothering alone	"I really don't have any help," "My baby has no father."
Real-life worry	"I don't have any money," "Will my baby be OK?" "It's not safe here."

The Quantitative-Qualitative "Debate"

Given their differences, it may come as no surprise that quantitative and qualitative research in psychology and related fields do not coexist in complete harmony. Some quantitative researchers

criticize qualitative methods on the grounds that they lack objectivity, are difficult to evaluate in terms of reliability and validity, and do not allow generalizations to people or situations other than those actually studied. At the same time, some qualitative researchers criticize quantitative methods on the grounds that they overlook the richness of human behaviour and experience and instead answer simple questions about easily quantifiable variables.

In general, however, qualitative researchers are well aware of the issues of objectivity, reliability, validity, and generalizability. In fact, they have developed a number of frameworks for addressing these issues (which are beyond the scope of our discussion). And in general, quantitative researchers are well aware of the issue of oversimplification. They do not believe that all human behaviour and experience can be adequately described in terms of a small number of variables and the statistical relationships among them. Instead, they use simplification as a strategy for uncovering general principles of human behaviour.

Many researchers from both the quantitative and qualitative camps now agree that the two approaches can and should be combined into what has come to be called mixed-methods research (Todd et al., 2004). (In fact, the studies by Lindqvist and colleagues and by Abrams and Curran both combined quantitative and qualitative approaches.) One approach to combining quantitative and qualitative research is to use qualitative research for hypothesis generation and quantitative research for hypothesis testing. Again, while a qualitative study might suggest that families who experience an unexpected suicide have more difficulty resolving the question of why, a well-designed quantitative study could test a hypothesis by measuring these specific variables in a large sample. A second approach to combining quantitative and qualitative research is referred to as triangulation. The idea is to use both quantitative and qualitative methods simultaneously to study the same general questions and compare the results. If the results of the quantitative and qualitative methods converge on the same general conclusion, they reinforce and enrich each other. If the results diverge, then

they suggest an interesting new question: Why do the results diverge and how can they be reconciled?

Using qualitative research can often help clarify quantitative results via triangulation. Trenor and her colleagues (2008) investigated the experience of female engineering students at a university. In the first phase, female engineering students were asked to complete a survey, where they rated a number of their perceptions, including their sense of belonging. Their results were compared across the student ethnicities, and statistically, the various ethnic groups showed no differences in their ratings of their sense of belonging. One might look at that result and conclude that ethnicity does not have anything to do with one's sense of belonging. However, in the second phase, the authors also conducted interviews with the students, and in those interviews, many minority students reported how the diversity of cultures at the university enhanced their sense of belonging. Without the qualitative component, we might have drawn the wrong conclusion about the quantitative results.

This example shows how qualitative and quantitative research work together to help us understand human behaviour. Some researchers have characterized qualitative research as best for identifying behaviours or the phenomenon whereas quantitative research is best for understanding meaning or identifying the mechanism. However, Bryman (2012) argues for breaking down the divide between these arbitrarily different ways of investigating the same questions.

Observational Research

The term **observational research** is used to refer to several different types of non-experimental studies in which behaviour is systematically observed and recorded. The goal of observational research is to describe a variable or set of variables. More generally, the goal is to obtain a snapshot of specific characteristics of an individual, group, or setting. As described previously, observational research is non-experimental because nothing is manipulated or controlled, and as such, we cannot arrive at causal conclusions using this approach. The data collected in observational research studies are often qualitative in nature but may also be quantitative or both (mixed-methods). There are several different types of observational methods that will be described below.

Naturalistic Observation

Naturalistic observation is an observational method that involves observing people's behaviour in the environment in which it typically occurs. Thus, naturalistic observation is a type of field research (as opposed to a type of laboratory research). Jane Goodall's famous research on chimpanzees is a classic example of naturalistic observation. Dr. Goodall spent three decades observing chimpanzees in their natural environment in East Africa. She examined such things as chimpanzees' social structure, mating patterns, gender roles, family structure, and care of offspring by observing them in the wild. However, naturalistic observation could more simply involve observing shoppers in a grocery store, children on a school playground, or psychiatric inpatients in their wards.

Researchers engaged in naturalistic observation usually make their observations as unobtrusively as possible so that participants are not aware that they are being studied. Such an approach is

called **disguised naturalistic observation.** Ethically, this method is considered to be acceptable if the participants remain anonymous and the behaviour occurs in a public setting where people would not normally have an expectation of privacy. Grocery shoppers putting items into their shopping carts, for example, are engaged in public behaviour that is easily observable by store employees and other shoppers. For this reason, most researchers would consider it ethically acceptable to observe them for a study. On the other hand, one of the arguments against the ethicality of the naturalistic observation of "bathroom behaviour" discussed earlier in the book is that people have a reasonable expectation of privacy in a restroom, even if it is public, and this expectation was violated.

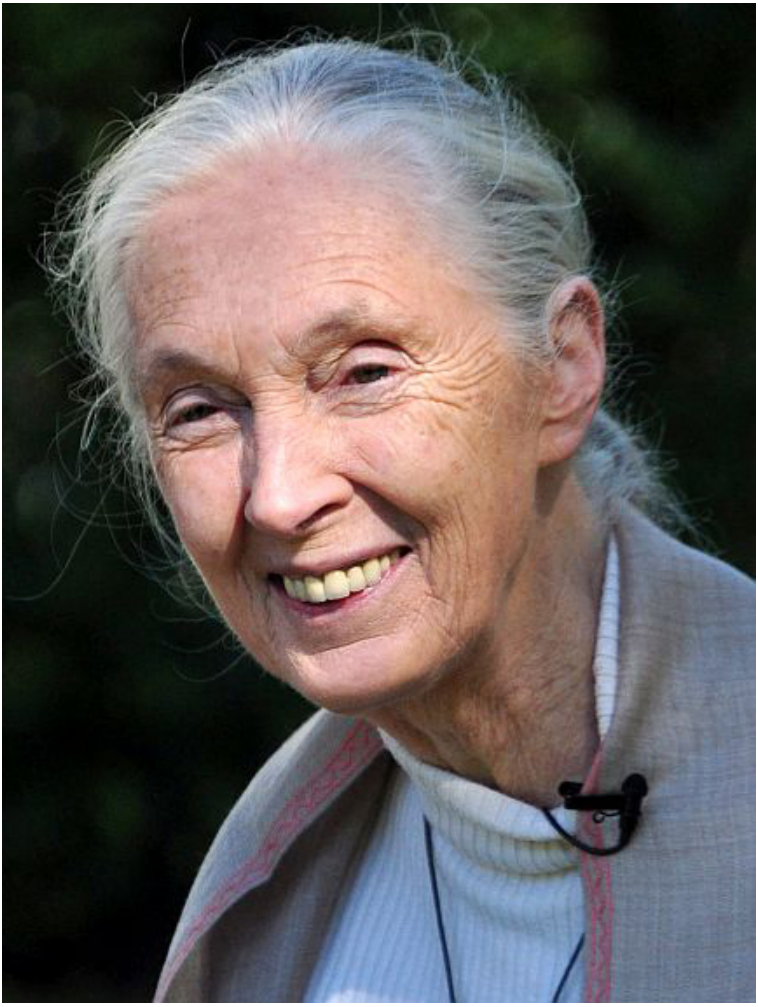


Figure 6.1 *Jane Goodall (Nick Step/Flickr)* CC BY 2.0

In cases where it is not ethical or practical to conduct disguised naturalistic observation, researchers can conduct **undisguised naturalistic observation**, where the participants are made aware of the researchers' presence and monitoring of their behaviour.

However, one concern with undisguised naturalistic observation is reactivity. **Reactivity** refers to when a measure changes participants' behaviour. In the case of undisguised naturalistic observation, the concern with reactivity is that when people know they are being observed and studied, they may act differently than they normally would. This type of reactivity is known as the **Hawthorne effect**. For instance, you may act much differently in a bar if you know that someone is observing you and recording your behaviours, which would invalidate the study. So disguised observation is less reactive and, therefore, can have higher validity because people are not aware that their behaviours are being observed and recorded.

However, we now know that people often become used to being observed, and with time, they begin to behave naturally in the researcher's presence. In other words, over time people habituate to being observed. Think about reality shows like Big Brother or Survivor where people are constantly being observed and recorded. While they may be on their best behaviour at first, in a fairly short amount of time, they are flirting, having sex, wearing next to nothing, screaming at each other, and occasionally behaving in ways that are embarrassing.

Participant Observation

Another approach to data collection in observational research is participant observation. In **participant observation**, researchers become active participants in the group or situation they are studying. Participant observation is very similar to naturalistic observation in that it involves observing people's behaviour in the environment in which it typically occurs. As with naturalistic observation, the data collected can include interviews (usually unstructured), notes based on their observations and interactions, documents, photographs, and other artifacts.

The only difference between naturalistic observation and participant observation is that researchers engaged in participant observation become active members of the group or situations they are studying. The basic rationale for participant observation is that there may be important information that is only accessible to, or can be interpreted only by, someone who is an active participant in the group or situation. Like naturalistic observation, participant observation can be either disguised or undisguised. In **disguised participant observation,** the researchers pretend to be members of the social group they are observing and conceal their true identity as researchers.

In a famous example of disguised participant observation, Leon Festinger and his colleagues (1956) infiltrated a doomsday cult known as the Seekers, whose members believed that the apocalypse would occur on December 21, 1954. Interested in studying how members of the group would cope psychologically when the prophecy inevitably failed, they carefully recorded the events and reactions of the cult members in the days before and after the supposed end of the world. Unsurprisingly, the cult members did not give up their belief but instead convinced themselves that it was their faith and efforts that saved the world from destruction. Festinger and his colleagues later published a book about this experience, which they used to illustrate the theory of cognitive dissonance.

In contrast with **undisguised participant observation,** the researchers become a part of the group they are studying and disclose their true identity as researchers to the group under investigation. Once again there are important ethical issues to consider with disguised participant observation. First, no informed consent can be obtained, and second, deception is being used. The researcher is deceiving the participants by intentionally withholding information about their motivations for being a part of the social group they are studying. But sometimes, disguised participation is the only way to access a protective group (like a cult). Further,

disguised participant observation is less prone to reactivity than undisguised participant observation.

Rosenhan's study (1973) of the experience of people in a psychiatric ward would be considered disguised participant observation because Rosenhan and his pseudopatients were admitted into psychiatric hospitals on the pretense of being patients so that they could observe the way that psychiatric patients are treated by staff. The staff and other patients were unaware of their true identities as researchers.

Another example of participant observation comes from a study by sociologist Amy Wilkins (2008) on a university-based religious organization that emphasized how happy its members were. Wilkins spent 12 months attending and participating in the group's meetings and social events and interviewed several group members. In her study, Wilkins identified several ways in which the group "enforced" happiness—for example, by continually talking about happiness, discouraging the expression of negative emotions, and using happiness as a way to distinguish themselves from other groups.

One of the primary benefits of participant observation is that the researchers are in a much better position to understand the viewpoint and experiences of the people they are studying when they are a part of the social group. The primary limitation with this approach is that the mere presence of the observer could affect the behaviour of the people being observed. While this is also a concern with naturalistic observation, additional concerns arise when researchers become active members of the social group they are studying because they may change the social dynamics and/or influence the behaviour of the people they are studying. Similarly, if the researcher acts as a participant observer, there can be concerns with biases resulting from developing relationships with the participants. Consequently, the researcher may become less objective, resulting in more experimenter bias.

Structured Observation

Another observational method is **structured observation**. Here, the investigator makes careful observations of one or more specific behaviours in a particular setting that is more structured than the settings used in naturalistic or participant observation. Often, the setting in which the observations are made is not the natural setting. Instead, the researcher may observe people in a laboratory environment. Alternatively, the researcher may observe people in a natural setting (like a classroom setting) that they have structured in some way, for instance by introducing some specific task for participants to engage in or by introducing a specific social situation or manipulation.

Structured observation is very similar to naturalistic observation and participant observation in that in all three cases researchers are observing naturally occurring behaviour; however, the emphasis in structured observation is on gathering quantitative rather than qualitative data. Researchers using this approach are interested in a limited set of behaviours. This allows them to quantify the behaviours they are observing. In other words, structured observation is less global than naturalistic or participant observation because the researcher engaged in structured observations is interested in a small number of specific behaviours. Therefore, rather than recording everything that happens, the researcher only focuses on very specific behaviours of interest.

A Structured Observation Example

Researchers Robert Levine and Ara Norenzayan (1999) used structured observation to study differences in the

"pace of life" across countries. One of their measures involved observing pedestrians in a large city to see how long it took them to walk 60 feet. They found that people in some countries walked reliably faster than people in other countries. For example, people in Canada and Sweden covered 60 feet in just under 13 seconds on average, while people in Brazil and Romania took close to 17 seconds.

When structured observation takes place in the complex and even chaotic "real world," the questions of when, where, and under what conditions the observations will be made, and who exactly will be observed are important to consider. Levine and Norenzayan described their sampling process as follows:

> *Male and female walking speed over a distance of 60 feet was measured in at least two locations in main downtown areas in each city. Measurements were taken during main business hours on clear summer days. All locations were flat, unobstructed, had broad sidewalks, and were sufficiently uncrowded to allow pedestrians to move at potentially maximum speeds. To control for the effects of socializing, only pedestrians walking alone were used. Children, individuals with obvious physical handicaps, and window-shoppers were not timed. Thirty-five men and 35 women were timed in most cities.* (p. 186)

Precise specification of the sampling process in this way makes data collection manageable for the observers and also provides some control over important extraneous variables. For example, by making their observations on clear summer days in all countries, Levine and Norenzayan (1999) controlled for effects of the weather

on people's walking speeds. In Levine and Norenzayan's study, measurement was relatively straightforward. They simply measured out a 60-foot distance along a city sidewalk and then used a stopwatch to time participants as they walked over that distance.

As another example, researchers Robert Kraut and Robert Johnston (1979) wanted to study bowlers' reactions to their shots, both when they were facing the pins and then when they turned toward their companions. But what "reactions" should they observe? Based on previous research and their own pilot testing, Kraut and Johnston created a list of reactions that included "closed smile," "open smile," "laugh," "neutral face," "look down," "look away," and "face cover" (covering one's face with one's hands). The observers committed this list to memory and then practiced by coding the reactions of bowlers who had been videotaped. During the actual study, the observers spoke into an audio recorder, describing the reactions they observed. Among the most interesting results of this study was that bowlers rarely smiled while they still faced the pins. They were much more likely to smile after they turned toward their companions, suggesting that smiling is not purely an expression of happiness but also a form of social communication.

In yet another example (this one in a laboratory environment), Dov Cohen and his colleagues (1996) had observers rate the emotional reactions of participants who had just been deliberately bumped into and insulted by a confederate after they dropped off a completed questionnaire at the end of a hallway. The confederate was posing as someone who worked in the same building and who was frustrated by having to close a file drawer twice in order to permit the participants to walk past them (first to drop off the questionnaire at the end of the hallway and once again on their way back to the room where they believed the study they signed up for was taking place). The two observers were positioned at different ends of the hallway so that they could read the participants' body language and hear anything they might say. Interestingly, the researchers hypothesized that participants from the southern

United States, which is one of several places in the world that has a "culture of honor," would react with more aggression than participants from the northern United States, a prediction that was, in fact, supported by the observational data.

When the observations require a judgment on the part of the observers—as in the studies by Kraut and Johnston and Cohen and his colleagues—a process referred to as coding is typically required. Coding generally requires clearly defining a set of target behaviours. The observers then categorize participants individually in terms of which behaviour they have engaged in and the number of times they engaged in each behaviour. The observers might even record the duration of each behaviour.

The target behaviours must be defined in such a way that guides different observers to code them in the same way. Researchers are expected to demonstrate the interrater reliability of their coding procedure by having multiple raters code the same behaviours independently and then showing that the different observers are in close agreement. Kraut and Johnston (1979), for example, video recorded a subset of their participants' reactions and had two observers independently code them. The two observers showed that they agreed on the reactions that were exhibited 97% of the time, indicating good interrater reliability.

One of the primary benefits of structured observation is that it is far more efficient than naturalistic and participant observation. Since the researchers are focused on specific behaviours, this reduces time and expense. Also, often times, the environment is structured to encourage the behaviours of interest, which again means that researchers do not have to invest as much time in waiting for the behaviours of interest to naturally occur.

Finally, researchers using this approach can clearly exert greater control over the environment. However, when researchers exert more control over the environment, it may make the environment less natural, which decreases external validity. It is less clear, for instance, whether structured observations made in a laboratory environment will generalize to a real-world environment.

Furthermore, since researchers engaged in structured observation are often not disguised there may be more concerns with reactivity.

Media Attributions

- **Figure 6.1** Jane Goodall by Nick Step, via Flickr, is used under a CC BY 2.0 license.

Case Studies

A **case study** is an in-depth examination of an individual. Sometimes, case studies are also completed on social units (e.g., a cult) and events (e.g., a natural disaster). Most commonly in psychology, however, case studies provide a detailed description and analysis of an individual. Often, the individual has a rare or unusual condition or disorder or has damage to a specific region of the brain.

Like many observational research methods, case studies tend to be more qualitative in nature. Case study methods involve an in-depth and often longitudinal examination of an individual. Depending on the focus of the case study, individuals may or may not be observed in their natural setting. If the natural setting is not what is of interest, then the individual may be brought into a therapist's office or researcher's lab for study. Also, the bulk of the case study report will focus on in-depth descriptions of the person rather than on statistical analyses. With that said, some quantitative data may also be included in the write-up of a case study. For instance, an individual's depression score may be compared to normative scores, or their score before and after treatment may be compared. As with other qualitative methods, a variety of different methods and tools can be used to collect information on the case. For instance, interviews, naturalistic observation, structured observation, psychological testing (e.g., IQ test), and/or physiological measurements (e.g., brain scans) may be used to collect information on an individual.

Patient H.M. is one of the most notorious case studies in psychology. H.M. suffered from intractable and very severe epilepsy. A surgeon localized H.M.'s epilepsy to his medial temporal lobe, and in 1953, he removed large sections of his hippocampus in an attempt to stop the seizures. The treatment was a success in that it resolved his epilepsy and his IQ and personality were unaffected.

However, the doctors soon realized that H.M. exhibited a strange

form of amnesia, called anterograde amnesia. H.M. was able to carry out a conversation and remember short strings of letters, digits, and words. Basically, his short-term memory was preserved. However, H.M. could not commit new events to memory. He lost the ability to transfer information from his short-term memory to his long-term memory, something memory researchers call consolidation. So, while he could carry on a conversation with someone, he would completely forget the conversation after it ended. This was an extremely important case study for memory researchers because it suggested that there is a dissociation between short-term memory and long-term memory and that these were two different abilities sub-served by different areas of the brain. It also suggested that the temporal lobes are particularly important for consolidating new information (i.e., for transferring information from short-term memory to long-term memory).

The history of psychology is filled with influential cases studies, such as Sigmund Freud's (1961) description of "Anna O." (see Note 6.1 "The Case of "Anna O.") and John Watson and Rosalie Rayner's (1920) description of Little Albert, who allegedly learned to fear a white rat—along with other furry objects—when the researchers repeatedly made a loud noise every time the rat approached him.

The Case of Anne O

Sigmund Freud (1961) used the case of a young woman he called "Anna O." to illustrate many principles of his theory of psychoanalysis. (Her real name was Bertha Pappenheim, and she was an early feminist who went on to make important contributions to the field of social work.) Anna had come to Freud's colleague Josef Breuer around 1880

with a variety of odd physical and psychological symptoms. One of them was that she was unable to drink any fluids for several weeks. According to Freud, she would take up the glass of water that she longed for, but as soon as it touched her lips, she would push it away like someone suffering from hydrophobia. "…She lived only on fruit, such as melons, etc., so as to lessen her tormenting thirst" (p. 9).

But according to Freud, a breakthrough came one day while Anna was under hypnosis.

> [S]he grumbled about her English "lady-companion," whom she did not care for, and went on to describe, with every sign of disgust, how she had once gone into this lady's room and how her little dog—horrid creature!—had drunk out of a glass there. The patient had said nothing, as she had wanted to be polite. After giving further energetic expression to the anger she had held back, she asked for something to drink, drank a large quantity of water without any difficulty, and awoke from her hypnosis with the glass at her lips; and thereupon the disturbance vanished, never to return. (p.9)

Freud's interpretation was that Anna had repressed the memory of this incident along with the emotion that it triggered and that this was what had caused her inability to drink. Furthermore, he believed that her recollection of the incident, along with her expression of the emotion she had repressed, caused the symptom to go away.

As an illustration of Freud's theory, the case study of Anna O. is quite effective. As evidence for the theory, however, it is essentially worthless. The description

provides no way of knowing whether Anna had really repressed the memory of the dog drinking from the glass, whether this repression had caused her inability to drink, or whether recalling this "trauma" relieved the symptom. It is also unclear from this case study how typical or atypical Anna's experience was.

Figure 6.2 *Anna O. "Anna O." was the subject of a famous case study used by Freud to illustrate the principles of psychoanalysis (Unidentified photographer/Wikimedia Commons)* Public Domain

Case studies are useful because they provide a level of detailed analysis not found in many other research methods and greater insights may be gained from this more detailed analysis. As a result of a case study, the researcher may gain a sharpened understanding of what might become important to look at more extensively in future more controlled research. Case studies are also often the only way to study rare conditions because it may be impossible to find a large enough sample of individuals with the condition to use quantitative methods. Although, at first glance, a case study of a rare individual might seem to tell us little about ourselves, they often provide insights into normal behaviour. The case of H.M. provided important insights into the role of the hippocampus in memory consolidation.

However, it is important to note that while case studies can provide *insights* into certain areas and variables to study and can be useful in helping develop theories, they should never be used as evidence for theories. In other words, case studies can be used as inspiration to formulate theories and hypotheses, but those hypotheses and theories then need to be formally tested using more rigorous quantitative methods. The reason case studies should not be used to provide support for theories is that they suffer from problems with both internal and external validity. Case studies lack the proper controls that true experiments contain. As such, they suffer from problems with internal validity, so they cannot be used to determine causation. For instance, during H.M.'s surgery, the surgeon may have accidentally lesioned another area of H.M.'s brain

(a possibility suggested by the dissection of H.M.'s brain following his death) and that lesion may have contributed to his inability to consolidate new information. The fact is that we cannot rule out these sorts of alternative explanations for case studies. So, as with all observational methods, case studies do not permit determination of causation.

In addition, because case studies are often of a single individual, and typically an abnormal individual, researchers cannot generalize their conclusions to other individuals. Recall that with most research designs, there is a trade-off between internal and external validity. With case studies, however, there are problems with both internal validity and external validity. So, there are limits both to the ability to determine causation and to generalize the results.

A final limitation of case studies is that ample opportunity exists for the theoretical biases of the researcher to colour or bias the case description. Indeed, there have been accusations that the woman who studied H.M. destroyed a lot of her data that were not published, and she has been called into question for destroying contradictory data that did not support her theory about how memories are consolidated. If you would like to learn more about Patient H.M.'s case, there is a fascinating New York Times article called "The Brain That Couldn't Remember" that describes some of the controversies that ensued after H.M.'s death and analysis of his brain (Dittrich, 2016).

Media Attributions

- **Figure 6.2** "Pappenheim 1882" by an unidentified photographer, via Wikimedia Commons, is in the public domain.

Archival Research

Another approach that is often considered observational research involves analyzing archival data that have already been collected for some other purpose. An example is a study by Brett Pelham and his colleagues (2005) on "implicit egotism"—the tendency for people to prefer people, places, and things that are similar to themselves. In one study, they examined Social Security records to show that women with the names Virginia, Georgia, Louise, and Florence were especially likely to have moved to the states of Virginia, Georgia, Louisiana, and Florida, respectively.

As with naturalistic observation, measurement can be more or less straightforward when working with archival data. For example, counting the number of people named Virginia who live in various states based on Social Security records is relatively straightforward. But consider a study by Christopher Peterson and his colleagues (1988) on the relationship between optimism and health using data that had been collected many years before for a study on adult development. In the 1940s, healthy male college students had completed an open-ended questionnaire about difficult wartime experiences. In the late 1980s, Peterson and his colleagues reviewed the men's questionnaire responses to obtain a measure of explanatory style—their habitual ways of explaining bad events that happen to them. More pessimistic people tend to blame themselves and expect long-term negative consequences that affect many aspects of their lives, while more optimistic people tend to blame outside forces and expect limited negative consequences.

To obtain a measure of explanatory style for each participant, the researchers used a procedure in which all negative events mentioned in the questionnaire responses and any causal explanations for them were identified and written on index cards. These were given to a separate group of raters who rated each explanation in terms of three separate dimensions of optimism-pessimism. These ratings were then averaged to produce an

explanatory style score for each participant. The researchers then assessed the statistical relationship between the men's explanatory style as undergraduate students and archival measures of their health at approximately 60 years of age. The primary result was that the more optimistic the men were as undergraduate students, the healthier they were as older men. Pearson's *r* was +.25.

This method is an example of **content analysis**—a family of systematic approaches to measurement using complex archival data. Just as structured observation requires specifying the behaviours of interest and then noting them as they occur, content analysis requires specifying keywords, phrases, or ideas and then finding all occurrences of them in the data. These occurrences can then be counted, timed (e.g., the amount of time devoted to entertainment topics on the nightly news show), or analyzed in a variety of other ways.

Summary

Key Takeaways

- Qualitative research is an important alternative to quantitative research in psychology. It generally involves asking broader research questions, collecting more detailed data (e.g., interviews), and using non-statistical analyses.
- Many researchers conceptualize quantitative and qualitative research as complementary and advocate combining them. For example, qualitative research can be used to generate hypotheses and quantitative research can test them.
- There are several different approaches to observational research including naturalistic observation, participant observation, structured observation, case studies, and archival research.
- Naturalistic observation is used to observe people in their natural setting; participant observation involves becoming an active member of the group being observed; structured observation involves coding a small number of behaviours in a quantitative manner; case studies are typically used to collect in-depth information on a single individual; and archival research involves analyzing existing data.

Acknowledgements

This chapter was adapted from parts of the following OER:

- *Research Methods in Psychology, 4th edition* by Jhiangiani et al. (2019), via Kwantlen Polytechnic University, is used under a CC BY-NC-SA 4.0 license.

References

Abrams, L. S., & Curran, L. (2009). "And you're telling me not to stress?" A grounded theory study of postpartum depression symptoms among low-income mothers. *Psychology of Women Quarterly*, 33(3), 351–362. https://doi.org/10.1177/0361684309033003 09

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An "experimental ethnography." *Journal of Personality and Social Psychology*, 70(5), 945-960. https://doi.org/10.1037//0022-3514.70.5.945

Dittrich, L. (2016, August 3). *The brain that couldn't remember*. The New York Times Magazine. https://www.nytimes.com/2016/08/07/magazine/the-brain-that-couldnt-remember.html?_r=0

Festinger, L., Riecken, H., & Schachter, S. (1956). *When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world*. University of Minnesota Press.

Freud, S. (1961). *Five lectures on psycho-analysis*. W. W. Norton & Company.

Geertz, C. (1973). *The interpretation of cultures*. Basic Books.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, 37(9), 1539–1553. https://doi.org/10.1037/0022-3514.37.9.1539

Levine, R. V., & Norenzayan, A. (1999). The pace of life in 31 countries. *Journal of Cross-Cultural Psychology*, 30(2), 178–205. https://doi.org/10.1177/0022022199030002003

Lindqvist, P., Johansson, L., & Karlsson, U. (2008). In the aftermath of teenage suicide: A qualitative study of the psychosocial

consequences for the surviving family members. *BMC Psychiatry*, 8, 26. https://doi.org/10.1186/1471-244X-8-26

Pelham, B. W., Carvallo, M., & Jones, J. T. (2005). Implicit egotism. *Current Directions in Psychological Science*, 14(2), 106–110. https://doi.org/10.1111/j.0963-7214.2005.00344.x

Peterson, C., Seligman, M. E. P., & Vaillant, G. E. (1988). Pessimistic explanatory style is a risk factor for physical illness: A thirty-five-year longitudinal study. *Journal of Personality and Social Psychology*, 55(1), 23–27. https://doi.org/10.1037/0022-3514.55.1.23

Rosenhan, D. L. (1973). On being sane in insane places. *Science*, 179(4070), 250–258. https://doi.org/10.1126/science.179.4070.250

Todd, Z., Nerlich, B., McKeown, S., & Clarke, D. D. (2004). *Mixing methods in psychology: The integration of qualitative and quantitative methods in theory and practice*. Psychology Press.

Trenor, J. M., Yu, S. L., Waight, C. L., Zerda, K. S., & Sha, T-L. (2008). The relations of ethnicity to female engineering students' educational experiences and college and career plans in an ethnically diverse learning environment. *Journal of Engineering Education*, 97(4), 449-465. https://doi.org/10.1002/j.2168-9830.2008.tb00992.x

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, 3(1), 1–14. https://doi.org/10.1037/h0069608

Wilkins, A. (2008). "Happier than Non-Christians": Collective emotions and symbolic boundaries among evangelical Christians. *Social Psychology Quarterly*, 71(3), 281–301. https://doi.org/10.1177/019027250807100308

VII. CORRELATIONS

Learning Objectives

- Define non-experimental research, distinguish it clearly from experimental research, and give several examples.
- Explain when a researcher might choose to conduct non-experimental research as opposed to experimental research.
- Define correlational research and give several examples.
- Explain why a researcher might choose to conduct correlational research rather than experimental research or another type of non-experimental research.
- Interpret the strength and direction of different correlation coefficients.
- Explain why correlation does not imply causation.

Introduction

What Is Non-Experimental Research?

Non-experimental research is research that lacks the manipulation of an independent variable. Rather than manipulating an independent variable, researchers conducting non-experimental research simply measure variables as they naturally occur (in the lab or real world).

Most researchers in psychology consider the distinction between experimental and non-experimental research to be an extremely important one. This is because although experimental research can provide strong evidence that changes in an independent variable cause differences in a dependent variable, non-experimental research generally cannot. As we will see, however, this inability to make causal conclusions does not mean that non-experimental research is less important than experimental research. It is simply used in cases where experimental research is not able to be carried out.

When to Use Non-Experimental Research

Experimental research is appropriate when the researcher has a specific research question or hypothesis about a causal relationship between two variables—and it is possible, feasible, and ethical to manipulate the independent variable. It stands to reason, therefore, that non-experimental research is appropriate—even necessary—when these conditions are not met. There are many times in which non-experimental research is preferred, including when:

- The research question or hypothesis relates to a single variable rather than a statistical relationship between two variables (e.g., how accurate are people's first impressions?).
- The research question pertains to a non-causal statistical relationship between variables (e.g., is there a correlation between verbal intelligence and mathematical intelligence?).
- The research question is about a causal relationship, but the independent variable cannot be manipulated or participants cannot be randomly assigned to conditions or orders of conditions for practical or ethical reasons (e.g., does damage to a person's hippocampus impair the formation of long-term memory traces?).
- The research question is broad and exploratory, or is about what it is like to have a particular experience (e.g., what is it like to be a working mother diagnosed with depression?).

Again, the choice between the experimental and non-experimental approaches is generally dictated by the nature of the research question. The four goals of science are to describe, to explain, to predict, and to apply. If the goal is to explain and the research question pertains to causal relationships, then the experimental approach is typically preferred. If the goal is to describe or predict, a non-experimental approach is appropriate.

But the two approaches can also be used to address the same research question in complementary ways. For example, in Milgram's (1974) original (non-experimental) obedience study, he was primarily interested in one variable—the extent to which participants obeyed the researcher when he told them to shock the confederate—and he observed all participants performing the same task under the same conditions. However, Milgram subsequently conducted experiments to explore the factors that affect obedience. He manipulated several independent variables, such as the distance between the experimenter and the participant, the participant and the confederate, and the location of the study.

Correlations as Non-Experimental Research

Non-experimental research falls into two broad categories: correlational research and observational research. The most common type of non-experimental research conducted in psychology is correlational research.

Correlational research is considered non-experimental because it focuses on the statistical relationship between two variables but does not include the manipulation of an independent variable. More specifically, in correlational research, the researcher measures two variables with little or no attempt to control extraneous variables and then assesses the relationship between them. As an example, a researcher interested in the relationship between self-esteem and school achievement could collect data on students' self-esteem and their GPAs to see if the two variables are statistically related.

Observational research (as discussed in the previous chapter) is non-experimental because it focuses on making observations of behaviour in a natural or laboratory setting without manipulating anything. Milgram's (1974) original obedience study was non-experimental in this way. He was primarily interested in the extent to which participants obeyed the researcher when he told them to shock the confederate, and he observed all participants performing the same task under the same conditions. The study by Loftus and Pickrell is also a good example of observational research. The variable was whether participants "remembered" having experienced mildly traumatic childhood events (e.g., getting lost in a shopping mall) that they had not actually experienced but that the researchers asked them about repeatedly. In this particular study, nearly a third of the participants "remembered" at least one event. (As with Milgram's original study, this study inspired several later experiments on the factors that affect false memories).

Cross-Sectional, Longitudinal, and Cross-Sequential Studies

When psychologists wish to study change over time (e.g., when developmental psychologists wish to study aging), they usually take one of three non-experimental approaches: cross-sectional, longitudinal, or cross-sequential. Cross-sectional studies involve comparing two or more pre-existing groups of people (e.g., children at different stages of development). What makes this approach non-experimental is that there is no manipulation of an independent variable and no random assignment of participants to groups.

Using this design, developmental psychologists compare groups of people of different ages (e.g., young adults spanning from 18-25 years of age versus older adults spanning 60-75 years of age) on various dependent variables (e.g., memory, depression, life satisfaction). Of course, the primary limitation of using this design to study the effects of aging is that differences between the groups other than age may account for differences in the dependent variable. For instance, differences between the groups may reflect the generation that people come from (a cohort effect) rather than a direct effect of age.

For this reason, longitudinal studies, in which one group of people is followed over time as they age, offer a superior means of studying the effects of aging. However, longitudinal studies are by definition more time consuming and so require a much greater investment on the part of the researcher and the participants.

A third approach, known as cross-sequential studies, combines elements of both cross-sectional and longitudinal studies. Rather than measuring differences between people in different age groups or following the same people over a long period of time, researchers adopting this approach choose a smaller period of time during which they follow people in different age groups. For example, they might measure changes over a 10-year period among participants who at the start of the study fall into the following age groups: 20 years old, 30 years old, 40 years old, 50 years old, and 60 years old. This design is advantageous because the researcher reaps the immediate benefits of being able to compare the age groups after the first assessment. Further, by following the different age groups over time they can subsequently determine whether the original differences they found across the age groups are due to true age effects or cohort effects.

Many observational research studies are more qualitative in nature. In qualitative research, the data are usually nonnumerical and, therefore, cannot be analyzed using statistical techniques. Rosenhan's (1973) observational study of the experience of people in psychiatric wards was primarily qualitative. The data were the notes taken by the "pseudo-patients"—the people pretending to have heard voices—along with their hospital records. Rosenhan's analysis consists mainly of a written description of the experiences of the pseudo-patients, supported by several concrete examples. To illustrate the hospital staff's tendency to "depersonalize" their patients, he noted, "Upon being admitted, I and other pseudo-patients took the initial physical examinations in a semi-public room, where staff members went about their own business as if we were not there" (Rosenhan, 1973, p. 256). Qualitative data has a separate set of analysis tools depending on the research question.

For example, thematic analysis would focus on themes that emerge in the data, or conversation analysis would focus on the way the words were said in an interview or focus group.

Internal Validity

Internal validity is the extent to which the design of a study supports the conclusion that changes in the independent variable caused any observed differences in the dependent variable. Figure 7.1 shows how experimental, quasi-experimental, and non-experimental (correlational) research vary in terms of internal validity. Experimental research tends to be highest in internal validity because the use of manipulation (of the independent variable) and control (of extraneous variables) help to rule out alternative explanations for the observed relationships. If the average score on the dependent variable in an experiment differs across conditions, it is quite likely that the independent variable is responsible for that difference. Non-experimental (correlational) research is lowest in internal validity because these designs fail to use manipulation or control.

Quasi-experimental research (which will be described in more detail in a subsequent chapter) falls in the middle because it contains some, but not all, of the features of a true experiment. For instance, it may fail to use random assignment to assign participants to groups or fail to use counterbalancing to control for potential order effects. For example, imagine that a researcher finds two similar schools, starts an anti-bullying program in one, and then finds fewer bullying incidents in that "treatment school" than in the "control school." While a comparison is being made with a control condition, the inability to randomly assign children to schools could still mean that students in the treatment school differed from students in the control school in some other way that could explain the difference in bullying (e.g., there may be a selection effect).

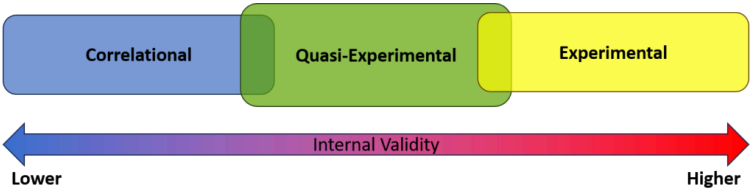


Figure 7.1 *Internal Validity of Correlation, Quasi-Experimental, and Experimental Studies. Experiments are generally high in internal validity, quasi-experiments lower, and correlation (non-experimental) studies lower still. (Jhangiani et al., 2019)* CC BY 4.0

Notice also in Figure 7.1 that there is some overlap in the internal validity of experiments, quasi-experiments, and correlational (non-experimental) studies. For example, a poorly designed experiment that includes many confounding variables can be lower in internal validity than a well-designed quasi-experiment with no obvious confounding variables. Internal validity is also only one of several validities that one might consider.

Media Attributions

- **Figure 7.1** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

Correlational Research

Correlational research is a type of non-experimental research in which the researcher measures two variables (binary or continuous) and assesses the statistical relationship (i.e., the correlation) between them with little or no effort to control extraneous variables. There are many reasons that researchers interested in statistical relationships between variables would choose to conduct a correlational study rather than an experiment.

The first is that they do not believe that the statistical relationship is a causal one or are not interested in causal relationships. Recall that two goals of science are to describe and to predict, and the correlational research strategy allows researchers to achieve both of these goals. Specifically, this strategy can be used to describe the strength and direction of the relationship between two variables and if there is a relationship between the variables, then the researchers can use scores on one variable to predict scores on the other.

Another reason that researchers would choose to use a correlational study rather than an experiment is that the statistical relationship of interest is thought to be causal, but the researcher cannot manipulate the independent variable because it is impossible, impractical, or unethical. For example, while a researcher might be interested in the relationship between the frequency people use cannabis and their memory abilities, they cannot ethically manipulate the frequency that people use cannabis. As such, they must rely on the correlational research strategy; they must simply measure the frequency that people use cannabis, measure their memory abilities using a standardized test of memory, and then determine whether the frequency people use cannabis is statistically related to memory test performance.

Correlation is also used to establish the reliability and validity of measurements. For example, a researcher might evaluate the validity of a brief extraversion test by administering it to a large

group of participants along with a longer extraversion test that has already been shown to be valid. This researcher might then check to see whether participants' scores on the brief test are strongly correlated with their scores on the longer one. Neither test score is thought to cause the other, so there is no independent variable to manipulate. In fact, the terms independent variable and dependent variable do not apply to this kind of research.

Another strength of correlational research is that it is often higher in external validity than experimental research. There is typically a trade-off between internal validity and external validity. As greater controls are added to experiments, internal validity is increased but often at the expense of external validity, as artificial conditions are introduced that do not exist in reality. In contrast, correlational studies typically have low internal validity because nothing is manipulated or controlled, but they often have high external validity. Since nothing is manipulated or controlled by the experimenter, the results are more likely to reflect relationships that exist in the real world.

Finally, extending upon this trade-off between internal and external validity, correlational research can help provide converging evidence for a theory. If a theory is supported by a true experiment that is high in internal validity as well as by a correlational study that is high in external validity, then the researchers can have more confidence in the validity of their theory. As a concrete example, correlational studies establishing that there is a relationship between watching violent television and aggressive behaviour have been complemented by experimental studies confirming that the relationship is a causal one (Bushman & Huesmann, 2001).

Does Correlational Research Always Involve Quantitative Variables?

A common misconception among beginning researchers is that

correlational research must involve two quantitative variables, such as scores on two extraversion tests or the number of daily hassles and number of symptoms people have experienced. However, the defining feature of correlational research is that the two variables are measured—neither one is manipulated—and this is true regardless of whether the variables are quantitative or categorical. Imagine, for example, that a researcher administers the Rosenberg Self-Esteem Scale to 50 American college students and 50 Japanese college students. Although this "feels" like a between-subjects experiment, it is a correlational study because the researcher did not manipulate the students' nationalities.

Figure 7.2 shows data from a hypothetical study on the relationship between people who make a daily list of things to do (a "to-do list") and their stress. Notice that it is unclear whether this is an experiment or a correlational study because it is unclear whether the independent variable was manipulated. If the researcher randomly assigned some participants to make daily to-do lists and others not to, then it is an experiment. If the researcher simply asked participants whether they made daily to-do lists, then it is a correlational study.

The distinction is important because if the study was an experiment, then it could be concluded that making the daily to-do lists reduced participants' stress. But if it was a correlational study, it could only be concluded that these variables are statistically related. Perhaps being stressed has a negative effect on people's ability to plan ahead (the directionality problem). Or perhaps people who are more conscientious are more likely to make to-do lists and less likely to be stressed (the third-variable problem). The crucial point is that what defines a study as experimental or correlational is not the variables being studied, nor whether the variables are quantitative or categorical, nor the type of graph or statistics used to analyze the data. What defines a study is how the study is conducted.

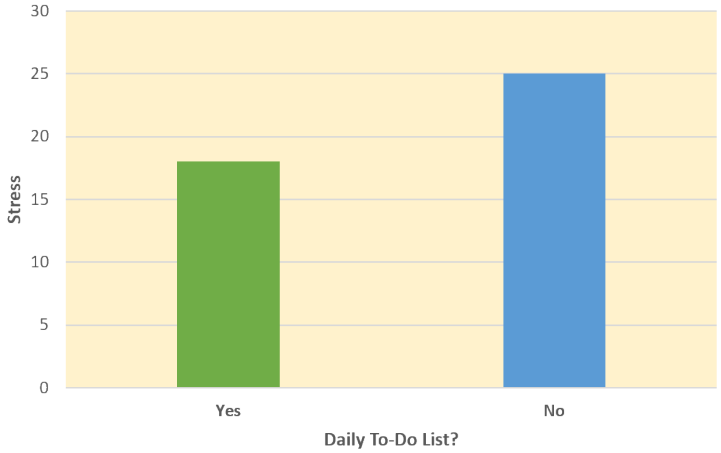


***Figure* 7.2** *Results of a hypothetical study on whether people who make daily to-do lists experience less stress than people who do not make such lists (Jhangiani et al., 2019)* CC BY 4.0

Data Collection in Correlational Research

Again, the defining feature of correlational research is that neither variable is manipulated. It does not matter how or where the variables are measured. For example, a researcher could have participants come to a laboratory to complete a computerized backward digit span task and a computerized risky decision-making task and then assess the relationship between participants' scores on the two tasks. Or a researcher could go to a shopping mall to ask people about their attitudes toward the environment and their shopping habits and then assess the relationship between these two variables. Both of these studies would be correlational because no independent variable is manipulated.

Correlations Between Quantitative Variables

Correlations between quantitative variables are often presented using **scatterplots**. Such graphs are a common way to illustrate correlations. However, it must be taken into account that while the dots may appear to show a convincing pattern suggesting a relationship between two measurements, a scatterplot is not a graph showing relationships. We will discuss more on this later.

Figure 7.3 shows some hypothetical data on the relationship between the amount of stress people are under and the number of physical symptoms they have. Each point in the scatterplot represents one person's score on both variables. For example, the circled point in Figure 7.3 represents a person whose stress score was 10 and who had three physical symptoms.

Taking all the points into account, one can see that people under more stress tend to have more physical symptoms. This is a good example of a **positive relationship**, in which higher scores on one variable tend to be associated with higher scores on the other. In other words, they move in the same direction, either both up or both down.

A **negative relationship** is one in which higher scores on one variable tend to be associated with lower scores on the other. In other words, they move in opposite directions. There is a negative relationship between stress and immune system functioning, for example, because higher stress is associated with lower immune system functioning.

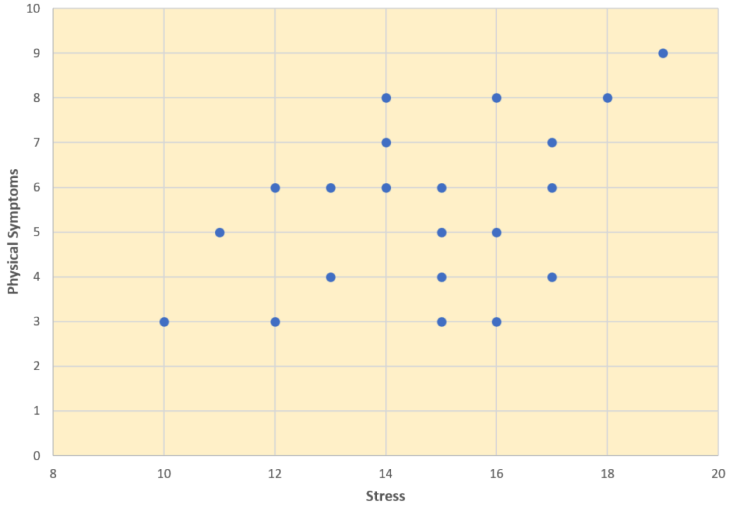


Figure 7.3 *Scatterplot showing a hypothetical positive relationship between stress and number of physical symptoms. The circled point represents a person whose stress score was 10 and who had three physical symptoms. Pearson's r for these data is +.51. (Jhangiani et al., 2019)* CC BY 4.0

The strength of a correlation between quantitative variables is typically measured using a statistic called **Pearson's Correlation Coefficient (or Pearson's *r*)**. As Figure 7.4 shows, Pearson's *r* ranges from −1.00 (the strongest possible negative relationship) to +1.00 (the strongest possible positive relationship). A value of 0 means there is no relationship between the two variables. When Pearson's *r* is 0, the points on a scatterplot form a shapeless "cloud." As its value moves toward −1.00 or +1.00, the points come closer and closer to falling on a single straight line. Correlation coefficients near ±.10 are considered small, values near ± .30 are considered medium, and values near ±.50 are considered large.

Notice that the sign of Pearson's *r* is unrelated to its strength. Pearson's *r* values of +.30 and −.30, for example, are equally strong; it is just that one represents a moderate positive relationship and the other a moderate negative relationship.

With the exception of reliability coefficients, most correlations that we find in psychology are small or moderate in size. The **Interpreting Correlations** website, created by Kristoffer Magnusson (n.d.), provides an excellent interactive visualization of correlations that permits you to adjust the strength and direction of a correlation while witnessing the corresponding changes to the scatterplot.

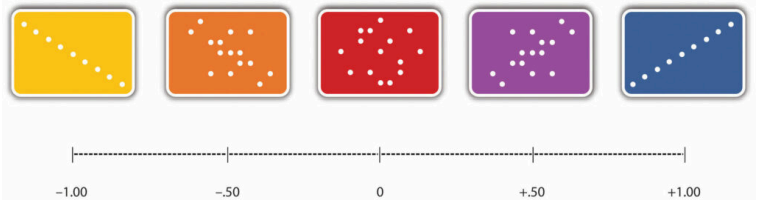


Figure 7.4 *Range of Pearson's r, from −1.00 (Strongest possible negative relationship), through 0 (No relationship), to +1.00 (Strongest possible positive relationship)* (*Jhangiani et al.*, 2019) CC BY 4.0

There are two common situations in which the value of Pearson's *r* can be misleading. Pearson's *r* is a good measure only for linear relationships, in which the points are best approximated by a straight line. It is not a good measure for nonlinear relationships, in which the points are better approximated by a curved line.

Figure 7.5, for example, shows a hypothetical relationship between the amount of sleep people get per night and their level of depression. In this example, the line that best approximates the points is a curve—a kind of upside-down "U"—because people who get about eight hours of sleep tend to be the least depressed. Those who get too little sleep and those who get too much sleep tend to be more depressed. Even though Figure 7.5 shows a fairly strong relationship between depression and sleep, Pearson's *r* would be close to zero because the points in the scatterplot are not well fit by a single straight line. This means that it is important to make a scatterplot and confirm that a relationship is approximately linear before using Pearson's *r*. Nonlinear relationships are fairly common

in psychology, but measuring their strength is beyond the scope of this book.

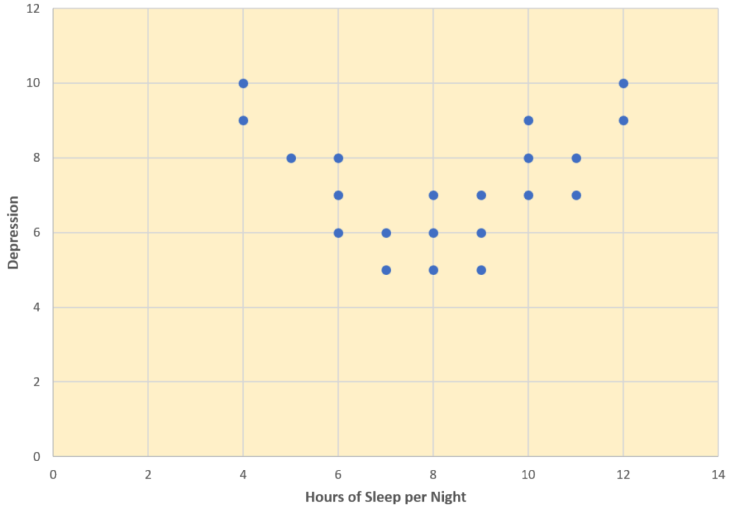


***Figure* 7.5** *Hypothetical nonlinear relationship between sleep and depression (Jhangiani et al., 2019)* CC BY 4.0

The other common situations in which the value of Pearson's *r* can be misleading is when one or both of the variables have a limited range in the sample relative to the population. This problem is referred to as **restriction of range**. Assume, for example, that there is a strong negative correlation between people's age and their enjoyment of hip-hop music, as shown by the scatterplot in Figure 7.6. Pearson's *r* here is −.77. However, if we were to collect data only from 18- to 24-year-olds—represented by the shaded area of Figure 7.6—then the relationship would seem to be quite weak. In fact, Pearson's *r* for this restricted range of ages is 0.

It is a good idea, therefore, to design studies that avoid restriction of range. For example, if age is one of your primary variables, then you can plan to collect data from people of a wide range of ages. Because restriction of range is not always anticipated or easily

avoidable, however, it is good practice to examine your data for possible restrictions of range and interpret Pearson's *r* in light of it. (There are also statistical methods to correct Pearson's *r* for restriction of range, but they are beyond the scope of this book).

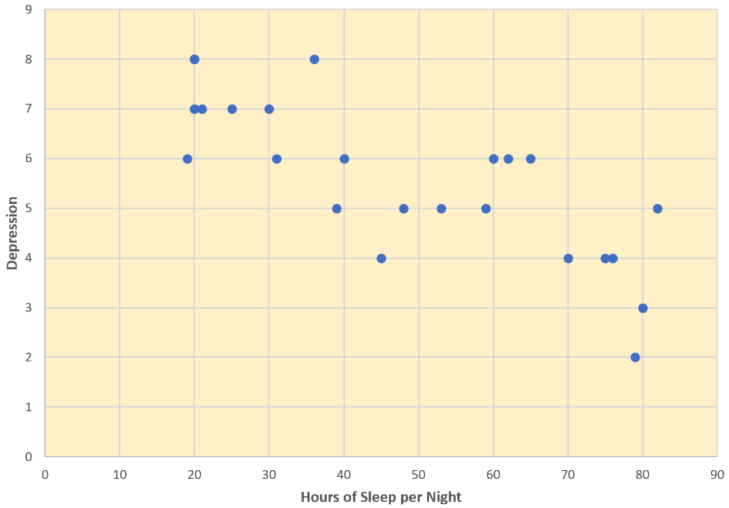


***Figure* 7.6** *Hypothetical data showing how a strong overall correlation can appear to be weak when one variable has a restricted range. The overall correlation here is –.77, but the correlation for the 18- to 24-year-olds is 0.* (*Jhangiani et al.*, 2019) CC BY 4.0

Media Attributions

- **Figure 7.2** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ
- **Figure 7.3** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

- **Figure 7.4** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ
- **Figure 7.5** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ
- **Figure 7.6** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

Jessica/Feb 20: One copyrighted image below

Correlation Does Not Imply Causation

You may have heard repeatedly that "Correlation does not imply causation." An amusing example of this comes from a 2012 study that showed a positive correlation (Pearson's $r = 0.79$) between the per capita chocolate consumption of a nation and the number of Nobel Prizes awarded to citizens of that nation (Messerli, 2012). It seems clear, however, that this does not mean that eating chocolate causes people to win Nobel Prizes, and it would not make sense to try to increase the number of Nobel Prizes won by recommending that parents feed their children more chocolate.

There are two reasons that correlation does not imply causation. The first is called the **directionality problem.** Two variables, X and Y, can be statistically related because X causes Y or because Y causes X. Consider, for example, a study showing that whether or not people exercise is statistically related to how happy they are—such that people who exercise are happier on average than people who do not. This statistical relationship is consistent with the idea that exercising causes happiness, but it is also consistent with the idea that happiness causes exercise. Perhaps being happy gives people more energy or leads them to seek opportunities to socialize with others by going to the gym.

The second reason that correlation does not imply causation is called the **third-variable problem.** Two variables, X and Y, can be statistically related not because X causes Y, or because Y causes X, but because some third variable, Z, causes both X and Y. For example, the fact that nations that have won more Nobel Prizes tend to have higher chocolate consumption probably reflects geography in that European countries tend to have higher rates of per capita chocolate consumption *and* invest more in education and technology (once again, per capita) than many other countries in the

world. Similarly, the statistical relationship between exercise and happiness could mean that some third variable, such as physical health, causes both of the others. Being physically healthy could cause people to exercise and cause them to be happier.

Correlations that are a result of a third-variable are often referred to as **spurious correlations**. Some excellent and amusing examples of spurious correlations can be found at Tyler Vigen's website (Figure 7.7 provides one such example).

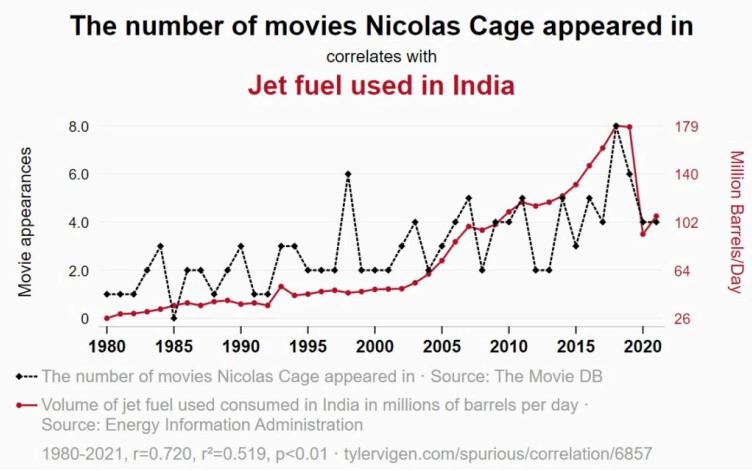


Figure 7.7 *Example of a spurious correlation. (Tyler Vigen/TylerVigen.com)* CC BY 4.0

5G Violence

The Covid-19 pandemic was a world-changing event in that it brought the entire world to a standstill for almost 2

years. There was a lot of uncertainty and fear about the origins of the virus as well as numerous conspiracy theories about causes and cures. One common image floating around the internet was similar to what we see in Figure 7.8: the overlap between 5G towers and the prevalence of COVID-19 cases.

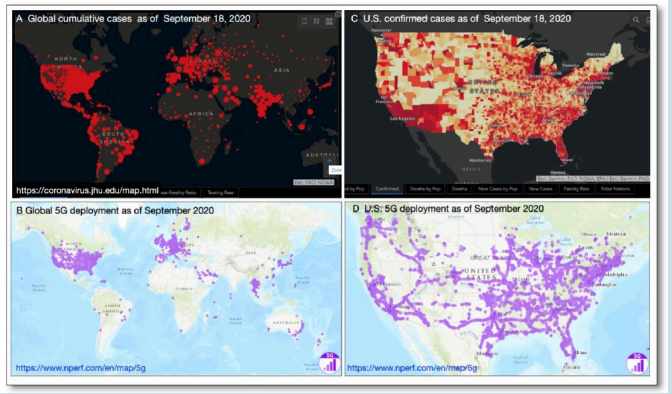


Figure 7.8 *Data for COVID-19 (as of September 18, 2020) and rollout of 5G as of September 2020 (Tsiang & Havas, 2021). [copyrighted image; need permission to use]*

As we can see, there is a strong correlation between 5G towers and COVID-19 cases. There was even a bombing in Nashville Tennessee linked to this conspiracy theory. However, what people did not understand was the third variable problem: correlation (however strong) does not mean causation. Is it possible that the distribution of 5G towers and the prevalence of an epidemic are correlation only because they are related to a third variable?

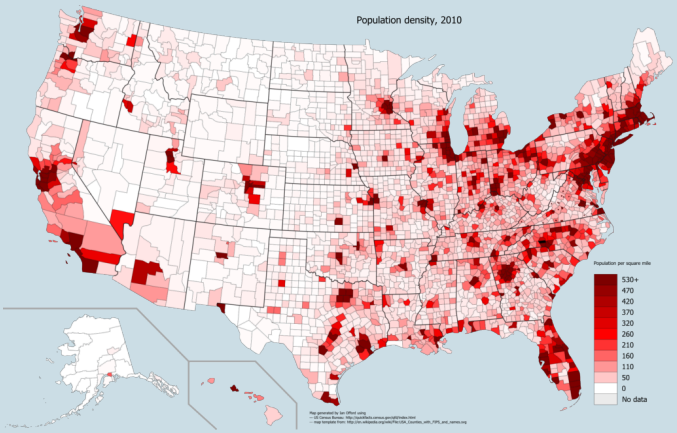


Figure 7.9 *Population density of the United States. From the 2006-2010 American Community Survey. (Wikimedia Commons, 2013)* CC–SA 3.0

As we can see from Figure 7.9, the population in the continental United States is not equally distributed across the land. Companies build 5G towers where they are needed. Therefore, they tend to build more towers in cities with high population density compared to rural areas. In the same vein, epidemics spread in places where human beings are more densely population and less so in rural areas with lower populations. This third variable (population density) is independently correlated with 5G towers and COVID-19 cases. However, if you just did a correlation between the latter two variables, you will find correlations. However, this does not mean there is a relationship between them.

Media Attributions

- **Figure 7.7** "The number of movies Nicolas Cage appeared in correlates with Jet fuel used in India" by Tyler Vigen, via TylerVigen.com, is used under a CC BY 4.0 license.
- **Figure 7.8** Tsiang, A. & Havas, M. (2021). COVID-19 Attributed Cases and Deaths are Statistically Higher in States and Counties with 5th Generation Millimeter Wave Wireless Telecommunications in the United States. *Medical Research Archives*, 9(4). https://doi.org/10.18103/mra.v9i4.2371
- **Figure 7.9** "United States Population Density" via Wikimedia Commons, is used under a CC–SA 3.0 license.

Summary

Key Takeaways

- Non-experimental research is research that lacks the manipulation of an independent variable.
- There are two broad types of non-experimental research: Correlational research focuses on statistical relationships between variables that are measured but not manipulated; and observational research in which participants are observed and their behaviour is recorded without the researcher interfering or manipulating any variables.
- In general, experimental research is high in internal validity, correlational research is low in internal validity, and quasi-experimental research is in between.
- Correlational research involves measuring two variables and assessing the relationship between them, with no manipulation of an independent variable.
- Correlation does not imply causation. A statistical relationship between two variables, X and Y, does not necessarily mean that X causes Y. It is also possible that Y causes X, or that a third variable, Z, causes both X and Y.
- While correlational research cannot be used to establish causal relationships between variables, correlational research does allow researchers to

achieve many other important objectives (establishing reliability and validity, providing converging evidence, describing relationships, and making predictions)

- Correlation coefficients can range from -1 to +1. The sign indicates the direction of the relationship between the variables and the numerical value indicates the strength of the relationship.

Jessica/Feb 20: Acknowledgements section Missing

References

Bushman, B. J., & Huesmann, L. R. (2001). Effects of televised violence on aggression. In D. G. Singer & J. L. Singer (Eds.), *Handbook of children and the media* (pp. 223–254). Sage Publications.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Diener, E. (2000). Subjective well-being: The science of happiness, and a proposal for a national index. *American Psychologist*, *55*(1), 34–43. https://doi.org/10.1037/0003-066X.55.1.34

Jouriles, E. N., Garrido, E., Rosenfield, D., & McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress. *Child Abuse & Neglect*, *33*(7), 451–460. https://doi.org/10.1016/j.chiabu.2008.11.005

Messerli, F. H. (2012). Chocolate consumption, cognitive function, and Nobel laureates. *New England Journal of Medicine*, *367*(16), 1562–1564. https://doi.org/10.1056/nejmon1211064

Milgram, S. (1974). *Obedience to authority: An experimental view*. Harper & Row.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral genetics* (5th ed.). Worth Publishers.

Radcliffe, N. M., & Klein, W. M. P. (2002). Dispositional, unrealistic, and comparative optimism: Differential relations with knowledge and processing of risk information and beliefs about personal risk. *Personality and Social Psychology Bulletin*, *28*(6), 836–846. https://doi.org/10.1177/0146167202289012

Rentfrow, P. J., & Gosling, S. D. (2008). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236–1256. https://doi.org/10.1037/0022-3514.84.6.1236

Rosenhan, D. L. (1973). On being sane in insane places. *Science*, 179(4070), 250–258. https://doi.org/10.1126/science.179.4070.250

Tsiang, A. & Havas, M. (2021). COVID-19 attributed cases and deaths are statistically higher in states and counties with 5th generation millimeter wave wireless telecommunications in the United States. *Medical Research Archives*, 9(4). https://doi.org/10.18103/mra.v9i4.2371

VIII. EXPERIMENTS

Learning Objectives

- Explain what an experiment is and recognize examples of studies that are experiments and ones that are not.
- Distinguish between the manipulation of the independent variable and control of extraneous variables and then explain the importance of each.
- Recognize examples of confounding variables and explain how they affect the internal validity of a study.
- Define what a control condition is, explain its purpose in research on treatment effectiveness, and describe some alternative types of control conditions.
- Explain the difference between between-subjects and within-subjects experiments, list some of the pros and cons of each approach, and decide which approach to use to answer a particular research question.
- Define random assignment, distinguish it from random sampling, explain its purpose in experimental research, and use some simple strategies to implement it
- Define several types of carryover effect, give examples of each, and explain how counterbalancing helps to deal with them.

- Explain what internal validity is and why experiments are considered to be high in internal validity.
- Explain what external validity is and evaluate studies in terms of their external validity.
- Explain the concepts of construct and statistical validity.
- Describe several strategies for recruiting participants for an experiment.
- Explain why it is important to standardize the procedure of an experiment and several ways to do this.
- Explain what pilot testing is and why it is important.

"Scientists have found…" is a statement heard throughout the internet. It is a statement that lends the writer (or speaker) some degree of legitimacy in saying what they are about to say. It is meant to convey that the statements are facts and that we can trust the findings. However, scientific experiments, like all human endeavours, have advantages and disadvantages. Understanding what scientists do when they perform experiments will help us better understand the definition of experiments, the way in which experiments are designed, and how to interpret the results of experiments.

In this chapter, we look at experiments in detail. We will first consider what sets experiments apart from other kinds of studies and why they support causal conclusions while other kinds of studies do not. We then look at two basic ways of designing an experiment—between-subjects designs and within-subjects designs—and discuss their pros and cons. Finally, we consider

several important practical issues that arise when conducting experiments.

Experimental Research

◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇
◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇

The Master said, "Yu, shall I teach you what knowledge is? When you know a thing, to hold that you know it; and when you do not know a thing, to allow that you do not know it – this is knowledge."

The Analects of Confucius – 2:17 (translated by James Legge, 1893)

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

They also suggested the theory that this phenomenon occurs because each witness feels less responsible for helping—a process referred to as the "diffusion of responsibility" (Darley & Latané, 1968). Darley and Latané noted that their ideas were consistent with many real-world cases. For example, a New York woman named Catherine "Kitty" Genovese was assaulted and murdered while several witnesses evidently failed to help. But Darley and Latané also understood that such isolated cases did not provide convincing evidence for their hypothesized "bystander effect." There was no way to know, for example, whether any of the witnesses to Kitty Genovese's murder would have helped had there been fewer of them.

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

gonna die-er-er-I'm…gonna die-er-help-er-er-seizure-er- [chokes, then quiet]" (Darley & Latané, 1968, p. 379).

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

Exercises

The story of Kitty Genovese has been told and retold in numerous psychology textbooks. The standard version is that there were 38 witnesses to the crime, that all of them watched (or listened) for an extended period of time, and that none of them did anything to help. However, recent scholarship suggests that the standard story is inaccurate in many ways (Manning et al., 2007). For example, only six eyewitnesses testified at the trial, none of them was aware that they were witnessing a lethal assault, and there have been several reports of witnesses calling the police or even coming to the aid of Kitty Genovese.

Although the standard story inspired a long line of research on the bystander effect and the diffusion of

responsibility, it may also have directed researchers' and students' attention away from other equally interesting and important issues in the psychology of helping—including the conditions in which people do in fact respond collectively to emergency situations.

The research that Darley and Latané conducted was a particular kind of study called an experiment. Experiments are used to determine not only whether there is a meaningful relationship between two variables but also whether the relationship is a causal one that is supported by statistical analysis. For this reason, experiments are one of the most common and useful tools in a psychological researcher's toolbox.

What Is an Experiment?

As we saw earlier in the book, an experiment is a type of study designed specifically to answer the question of whether there is a causal relationship between two variables. In other words, whether changes in one variable (referred to as an independent variable) cause a change in another variable (referred to as a dependent variable). Experiments have two fundamental features. The first is that the researchers manipulate, or systematically vary, the level of the independent variable. The different levels of the independent variable are called conditions.

For example, in Darley and Latané's (1968) experiment, the independent variable was the number of witnesses that participants believed to be present. The researchers manipulated this independent variable by telling participants that there were either one, two, or five other students involved in the discussion, thereby creating three conditions. For a new researcher, it is easy to confuse these terms by believing there are three independent variables in this situation: one, two, or five students involved in the discussion, but there is actually only one independent variable (number of witnesses) with three different levels or conditions (one, two, or five students).

The second fundamental feature of an experiment is that the researcher exerts control over, or minimizes the variability in, variables other than the independent and dependent variable. These other variables are called extraneous variables. Darley and Latané (1968) tested all their participants in the same room, exposed them to the same emergency situation, and so on. They also randomly assigned their participants to conditions so that the three groups would be similar to each other to begin with.

Notice that although the words manipulation and control have similar meanings in everyday language, researchers make a clear distinction between them. They manipulate the independent

variable by systematically changing its levels and control other variables by holding them constant.

Manipulation of the Independent Variable

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

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

of the independent variable is crucial for eliminating potential alternative explanations for the results.

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

Independent variables can be manipulated to create two conditions, and experiments involving a single independent variable with two conditions are often referred to as a single-factor two-level design. However, sometimes greater insights can be gained by adding more conditions to an experiment. When an experiment has one independent variable that is manipulated to produce more than two conditions, it is referred to as a single-factor multi-level design. So rather than comparing a condition in which there was one witness to a condition in which there were five witnesses (which would represent a single-factor two-level design), Darley and Latané's (1968) experiment used a single factor multi-level design, by manipulating the independent variable to produce three conditions (one-witness, two-witnesses, and five-witnesses conditions).

Control of Extraneous Variables

As we have seen previously in the chapter, an extraneous variable is anything that varies in the context of a study other than the independent and dependent variables. In an experiment on the

effect of expressive writing on health, for example, extraneous variables would include participant variables (individual differences), such as their writing ability, their diet, and their gender. They would also include situational or task variables, such as the time of day when participants write, whether they write by hand or on a computer, and the weather.

Extraneous variables pose a problem because many of them are likely to have some effect on the dependent variable. For example, participants' health will be affected by many things other than whether or not they engage in expressive writing. This influencing factor can make it difficult to separate the effect of the independent variable from the effects of the extraneous variables, which is why it is important to control extraneous variables by holding them constant.

Extraneous Variables as "Noise"

Extraneous variables make it difficult to detect the effect of the independent variable in two ways. One is by adding variability or "noise" to the data. Imagine a simple experiment on the effect of mood (happy vs. sad) on the number of happy childhood events people are able to recall. Participants are put into a negative or positive mood (by showing them a happy or sad video clip) and then asked to recall as many happy childhood events as they can. The two leftmost columns of Table 8.1 show what the data might look like if there were no extraneous variables and if the number of happy childhood events participants recalled was affected only by their moods.

Every participant in the happy mood condition recalled exactly four happy childhood events, and every participant in the sad mood condition recalled exactly three. The effect of mood here is quite obvious. In reality, however, the data would probably look more like those in the two rightmost columns of Table 8.1. Even in the

happy mood condition, some participants would recall fewer happy memories because they have fewer to draw on, use less effective recall strategies, or are less motivated. And even in the sad mood condition, some participants would recall more happy childhood memories because they have more happy memories to draw on, use more effective recall strategies, or are more motivated.

Although the mean difference between the two groups is the same as in the idealized data, this difference is much less obvious in the context of the greater variability in the data. Thus, one reason researchers try to control extraneous variables is so that their data look more like the idealized data in Table 8.1, which makes the effect of the independent variable easier to detect (although real data never look quite that good).

Table 8.1 Hypothetical Noiseless Data and Realistic Noisy Data

Idealized "noiseless" data: Happy mood	Idealized "noiseless" data: Sad mood	Realistic "noisy" data: Happy mood	Realistic "noisy" data: Sad mood
4	3	3	1
4	3	6	3
4	3	2	4
4	3	4	0
4	3	5	5
4	3	2	7
4	3	3	2
4	3	1	5
4	3	6	1
4	3	8	2
M = 4	M = 3	M = 4	M = 3

One way to control extraneous variables is to hold them constant. This technique can mean holding situation or task variables constant by testing all participants in the same location, giving them

identical instructions, treating them in the same way, and so on. It can also mean holding participant variables constant. For example, many studies of language limit participants to right-handed people, who generally have their language areas isolated in their left cerebral hemispheres. Left-handed people are more likely to have their language areas isolated in their right cerebral hemispheres or distributed across both hemispheres, which can change the way they process language and, thereby, add noise to the data.

In principle, researchers can control extraneous variables by limiting participants to one very specific category of people—such as 20-year-old, heterosexual, female, right-handed psychology majors. The obvious downside to this approach is that it would lower the external validity of the study—in particular, the extent to which the results can be generalized beyond the people actually studied. For example, it might be unclear whether results obtained with a sample of younger lesbian women would apply to older gay men. In many situations, the advantages of a diverse sample (increased external validity) outweigh the reduction in noise achieved by a homogeneous one.

Extraneous Variables as Confounding Variables

The second way that extraneous variables can make it difficult to detect the effect of the independent variable is by becoming confounding variables. A confounding variable is an extraneous variable that differs on average across levels of the independent variable (i.e., an extraneous variable that varies systematically with the independent variable). For example, in almost all experiments, participants' intelligence quotients (IQs) will be an extraneous variable. But as long as there are participants with lower and higher IQs in each condition so that the average IQ is roughly equal across the conditions, then this variation is probably acceptable (and may even be desirable). What would be bad, however, would be for

participants in one condition to have substantially lower IQs on average and participants in another condition to have substantially higher IQs on average. In this case, IQ would be a confounding variable.

To confound means to confuse, and this effect is exactly why confounding variables are undesirable. Because they differ systematically across conditions—just like the independent variable—they provide an alternative explanation for any observed difference in the dependent variable.

Figure 8.1 shows the results of a hypothetical study, in which participants in a positive mood condition scored higher on a memory task than participants in a negative mood condition. But if IQ is a confounding variable—with participants in the positive mood condition having higher IQs on average than participants in the negative mood condition—then it is unclear whether it was the positive moods or the higher IQs that caused participants in the first condition to score higher.

One way to avoid confounding variables is by holding extraneous variables constant. For example, one could prevent IQ from becoming a confounding variable by limiting participants only to those with IQs of exactly 100. But this approach is not always desirable for reasons we have already discussed. A second and much more general approach—random assignment to conditions—will be discussed in detail shortly.

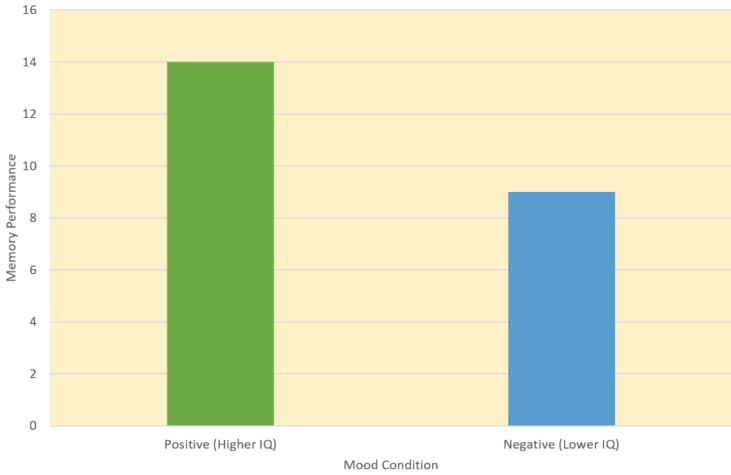


Figure 8.1 *Hypothetical results from a study on the effect of mood on memory. Because IQ also differs across conditions, it is a confounding variable (Jhangiani et al., 2019).* *Jessica/Feb 20: Image license missing*

Media Attributions

- **Figure 8.1** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

Treatment and Control Conditions

In psychological research, a treatment is any intervention meant to change people's behaviour for the better. This intervention includes psychotherapies and medical treatments for psychological disorders but also interventions designed to improve learning, promote conservation, reduce prejudice, and so on. To determine whether a treatment works, participants are randomly assigned to either a treatment condition, in which they receive the treatment, or a control condition, in which they do not receive the treatment. If participants in the treatment condition end up better off than participants in the control condition—for example, they are less depressed, learn faster, conserve more, or express less prejudice—then the researcher can conclude that the treatment works. In research on the effectiveness of psychotherapies and medical treatments, this type of experiment is often called a randomized clinical trial.

There are different types of control conditions. In a no-treatment control condition, participants receive no treatment whatsoever. One problem with this approach, however, is the existence of placebo effects. A placebo is a simulated treatment that lacks any active ingredient or element that should make it effective, and a placebo effect is a positive effect of such a treatment. Many folk remedies that seem to work—such as eating chicken soup for a cold or placing soap under the bed sheets to stop nighttime leg cramps—are probably nothing more than placebos. Although placebo effects are not well understood, they are probably driven primarily by people's expectations that they will improve. Having the expectation to improve can result in reduced stress, anxiety, and depression, which can alter perceptions and even improve immune system functioning (Price et al., 2008).

Placebo effects are interesting in their own right (see Note "The Powerful Placebo"), but they also pose a serious problem for researchers who want to determine whether a treatment works.

Figure 8.2 shows some hypothetical results in which participants in a treatment condition improved more on average than participants in a no-treatment control condition. If these conditions (the two leftmost bars in Figure 8.2) were the only conditions in this experiment, however, one could not conclude that the treatment worked. It could be instead that participants in the treatment group improved more because they expected to improve, while those in the no-treatment control condition did not.

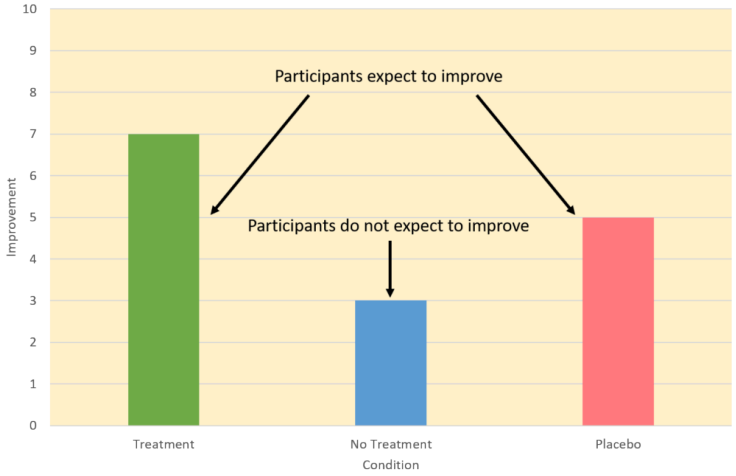


Figure 8.2 *Hypothetical results from a study including treatment, no-treatment, and placebo conditions (Jhangiani et al., 2019) Jessica/Feb 20: Image license missing*

Fortunately, there are several solutions to this problem. One is to include a placebo control condition, in which participants receive a placebo that looks much like the treatment but lacks the active ingredient or element thought to be responsible for the treatment's effectiveness. When participants in a treatment condition take a pill, for example, then those in a placebo control condition would take an identical-looking pill that lacks the active ingredient in the treatment (a "sugar pill"). In research on psychotherapy

effectiveness, the placebo might involve going to a psychotherapist and talking in an unstructured way about one's problems. The idea is that if participants in both the treatment and the placebo control groups expect to improve, then any improvement in the treatment group over and above that in the placebo control group must have been caused by the treatment and not by participants' expectations.

Of course, the principle of informed consent requires that participants be told that they will be assigned to either a treatment or a placebo control condition—even though they cannot be told which until the experiment ends. In many cases the participants who had been in the control condition are then offered an opportunity to have the real treatment.

An alternative approach is to use a wait-list control condition, in which participants are told that they will receive the treatment but must wait until the participants in the treatment condition have already received it. This disclosure allows researchers to compare participants who have received the treatment with participants who are not currently receiving it but who still expect to improve (eventually).

A final solution to the problem of placebo effects is to leave out the control condition completely and compare any new treatment with the best available alternative treatment. For example, a new treatment for simple phobia could be compared with standard exposure therapy. Because participants in both conditions receive a treatment, their expectations about improvement should be similar. This approach also makes sense because once there is an effective treatment, the interesting question about a new treatment is not simply "Does it work?" but "Does it work better than what is already available?

The Powerful Placebo

Many people are not surprised that placebos can have a positive effect on disorders that seem fundamentally psychological, including depression, anxiety, and insomnia. However, placebos can also have a positive effect on disorders that most people think of as fundamentally physiological. These include asthma, ulcers, and warts (Shapiro & Shapiro, 1999). There is even evidence that placebo surgery—also called "sham surgery"—can be as effective as actual surgery.

Medical researcher J. Bruce Moseley and his colleagues (2002) conducted a study on the effectiveness of two arthroscopic surgery procedures for osteoarthritis of the knee. The control participants in this study were prepped for surgery, received a tranquilizer, and even received three small incisions in their knees. But they did not receive the actual arthroscopic surgical procedure. Note that the Institutional Review Board would have carefully considered the use of deception in this case and judged that the benefits of using it outweighed the risks and that there was no other way to answer the research question (about the effectiveness of a placebo procedure) without it.

The surprising result was that all participants improved in terms of both knee pain and function, and the sham surgery group improved just as much as the treatment groups. According to the researchers, "This study provides strong evidence that arthroscopic lavage with or without debridement [the surgical procedures used] is not better than and appears to be equivalent to a placebo procedure in improving knee pain and self-reported function" (Moseley et al., 2002, p. 85).

Media Attributions

- **Figure 8.2** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

Experimental Design

In this section, we look at some different ways to design an experiment. The primary distinction we will make is between approaches in which each participant experiences one level of the independent variable and approaches in which each participant experiences all levels of the independent variable. The former are called between-subject experiments, and the latter are called within-subject experiments.

Between-Subjects Experiments

In a between-subjects experiment, each participant is tested in only one condition. For example, a researcher with a sample of 100 university students might assign half of them to write about a traumatic event and the other half write about a neutral event. Or a researcher with a sample of 60 people with severe agoraphobia (fear of open spaces) might assign 20 of them to receive each of three different treatments for that disorder.

It is essential in a between-subjects experiment that the researcher assigns participants to conditions so that the different groups are, on average, highly similar to each other. Those in a trauma condition and a neutral condition, for example, should include a similar proportion of men and women and should have similar average IQs, levels of motivation, numbers of health problems, and so on. This matching is a matter of controlling these extraneous participant variables across conditions so that they do not become confounding variables.

Random Assignment

The primary way that researchers accomplish this kind of control of extraneous variables across conditions is called random assignment, which means using a random process to decide which participants are tested in which conditions. Do not confuse random assignment with random sampling:

- **Random sampling** is a method for selecting a sample from a population, and it is rarely used in psychological research.
- **Random assignment** is a method for assigning participants in a sample to the different conditions, and it is an important element of all experimental research in psychology and other fields too.

In its strictest sense, random assignment should meet two criteria. One is that each participant has an equal chance of being assigned to each condition (e.g., a 50% chance of being assigned to each of two conditions). The second is that each participant is assigned to a condition independently of other participants. Thus, one way to assign participants to two conditions would be to flip a coin for each one. If the coin lands heads, the participant is assigned to Condition A, and if it lands tails, the participant is assigned to Condition B. For three conditions, one could use a computer to generate a random integer from 1 to 3 for each participant. If the integer is 1, the participant is assigned to Condition A; if it is 2, the participant is assigned to Condition B; and if it is 3, the participant is assigned to Condition C. In practice, a full sequence of conditions—one for each participant expected to be in the experiment—is usually created ahead of time, and each new participant is assigned to the next condition in the sequence as they are tested. When the procedure is computerized, the computer program often handles the random assignment.

One problem with coin flipping and other strict procedures for random assignment is that they are likely to result in unequal

sample sizes in the different conditions. Unequal sample sizes are generally not a serious problem, and you should never throw away data you have already collected to achieve equal sample sizes. However, for a fixed number of participants, it is statistically most efficient to divide them into equal-sized groups. It is standard practice, therefore, to use a kind of modified random assignment that keeps the number of participants in each group as similar as possible.

One approach is block randomization. In block randomization, all the conditions occur once in the sequence before any of them is repeated. Then, they all occur again before any of them is repeated again. Within each of these "blocks," the conditions occur in a random order. Again, the sequence of conditions is usually generated before any participants are tested, and each new participant is assigned to the next condition in the sequence. Table 8.2 shows such a sequence for assigning nine participants to three conditions. The Research Randomizer website will generate block randomization sequences for any number of participants and conditions (Urbaniak & Plous, n.d.). Again, when the procedure is computerized, the computer program often handles the block randomization.

Table 8.2 Block Randomization Sequence for Assigning Nine Participants to Three Conditions

Participant	Condition
1	A
2	C
3	B
4	B
5	C
6	A
7	C
8	B
9	A

Random assignment is not guaranteed to control all extraneous variables across conditions. The process is random, so it is always possible that just by chance, the participants in one condition might turn out to be substantially older, less tired, more motivated, or less depressed on average than the participants in another condition. However, there are some reasons that this possibility is not a major concern. One is that random assignment works better than one might expect, especially for large samples. Another is that the inferential statistics that researchers use to decide whether a difference between groups reflects a difference in the population takes the "fallibility" of random assignment into account. Yet another reason is that even if random assignment does result in a confounding variable and, therefore, produces misleading results, this confound is likely to be detected when the experiment is replicated. The upshot is that random assignment to conditions—although not infallible in terms of controlling extraneous variables—is always considered a strength of a research design.

Matched Groups

An alternative to simple random assignment of participants to conditions is the use of a matched-groups design. Using this design, participants in the various conditions are matched on the dependent variable or on some extraneous variable(s) prior the manipulation of the independent variable. This guarantees that these variables will not be confounded across the experimental conditions.

For instance, if we want to determine whether expressive writing affects people's health then we could start by measuring various health-related variables in our prospective research participants. We could then use that information to rank-order participants according to how healthy or unhealthy they are. Next, the two healthiest participants would be randomly assigned to complete different conditions (one would be randomly assigned to the traumatic experiences writing condition and the other to the neutral writing condition). The next two healthiest participants would then be randomly assigned to complete different conditions, and so on until the two least healthy participants. This method would ensure that participants in the traumatic experiences writing condition are matched to participants in the neutral writing condition with respect to health at the beginning of the study. If at the end of the experiment, a difference in health was detected across the two conditions, then we would know that it is due to the writing manipulation and not to pre-existing differences in health.

Within-Subjects Experiments

In a within-subjects experiment, each participant is tested under all conditions. Consider an experiment on the effect of a defendant's physical attractiveness on judgments of his guilt. Again, in a

between-subjects experiment, one group of participants would be shown an attractive defendant and asked to judge his guilt, and another group of participants would be shown an unattractive defendant and asked to judge his guilt. In a within-subjects experiment, however, the same group of participants would judge the guilt of both an attractive and an unattractive defendant.

The primary advantage of this approach is that it provides maximum control of extraneous participant variables. Participants in all conditions have the same mean IQ, same socioeconomic status, same number of siblings, and so on—because they are the very same people. Within-subjects experiments also make it possible to use statistical procedures that remove the effect of these extraneous participant variables on the dependent variable and, therefore, make the data less "noisy" and the effect of the independent variable easier to detect. We will look more closely at this idea later in the book. However, not all experiments can use a within-subjects design nor would it be desirable to do so.

In this section, we look at some different ways to design an experiment. The primary distinction we will make is between approaches in which each participant experiences one level of the independent variable and approaches in which each participant experiences all levels of the independent variable. The former are called between-subject experiments, and the latter are called within-subject experiments.

Carryover Effects and Counterbalancing

The primary disadvantage of within-subjects designs is that they can result in order effects. An order effect occurs when participants' responses in the various conditions are affected by the order of conditions to which they were exposed. One type of order effect is a carryover effect. A carryover effect is an effect of being tested in one condition on participants' behaviour in later conditions. One

type of carryover effect is a practice effect, where participants perform a task better in later conditions because they have had a chance to practice it. Another type is a fatigue effect, where participants perform a task worse in later conditions because they become tired or bored. Being tested in one condition can also change how participants perceive stimuli or interpret their task in later conditions. This type of effect is called a context effect (or contrast effect). For example, an average-looking defendant might be judged more harshly when participants have just judged an attractive defendant than when they have just judged an unattractive defendant.

Within-subjects experiments also make it easier for participants to guess the hypothesis. For example, a participant who is asked to judge the guilt of an attractive defendant and then is asked to judge the guilt of an unattractive defendant is likely to guess that the hypothesis is that defendant attractiveness affects judgments of guilt. This knowledge could lead the participant to judge the unattractive defendant more harshly because he thinks this is what he is expected to do. Or it could make participants judge the two defendants similarly in an effort to be "fair."

Carryover effects can be interesting in their own right. (Does the attractiveness of one person depend on the attractiveness of other people that we have seen recently?) But when they are not the focus of the research, carryover effects can be problematic. Imagine, for example, that participants judge the guilt of an attractive defendant and then judge the guilt of an unattractive defendant. If they judge the unattractive defendant more harshly, this might be because of his unattractiveness. But it could be instead that they judge him more harshly because they are becoming bored or tired. In other words, the order of the conditions is a confounding variable. The attractive condition is always the first condition and the unattractive condition the second. Thus, any difference between the conditions in terms of the dependent variable could be caused by the order of the conditions and not the independent variable itself.

There is a solution to the problem of order effects, however, that can be used in many situations. It is counterbalancing, which means testing different participants in different orders. The best method of counterbalancing is complete counterbalancing in which an equal number of participants complete each possible order of conditions. For example, half of the participants would be tested in the attractive defendant condition followed by the unattractive defendant condition, and the other half would be tested in the unattractive condition followed by the attractive condition. With three conditions, there would be six different orders (ABC, ACB, BAC, BCA, CAB, and CBA), so some participants would be tested in each of the six orders. With four conditions, there would be 24 different orders; with five conditions there would be 120 possible orders.

With counterbalancing, participants are assigned to orders randomly, using the techniques we have already discussed. Thus, random assignment plays an important role in within-subjects designs just as in between-subjects designs. Here, instead of randomly assigning to conditions, they are randomly assigned to different orders of conditions. In fact, it can safely be said that if a study does not involve random assignment in one form or another, it is not an experiment.

A more efficient way of counterbalancing is through a Latin square design, which randomizes through having equal rows and columns. For example, if you have four treatments, you must have four versions. Like a Sudoku puzzle, no treatment can repeat in a row or column. For four versions of four treatments, the Latin square design would look like:

Latin Square Design

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Jessica/Feb 20: Does this need to have a table number?

You can see in the diagram above that the square has been constructed to ensure that each condition appears at each ordinal position (A appears first once, second once, third once, and fourth once) and that each condition precedes and follows each other condition one time. A Latin square for an experiment with six conditions would by 6×6 in dimension, one for an experiment with right conditions would be 8×8 in dimension, and so on. So, while complete counterbalancing of six conditions would require 720 orders, a Latin square would only require six orders.

Finally, when the number of conditions is large, experiments can use random counterbalancing in which the order of the conditions is randomly determined for each participant. Using this technique, every possible order of conditions is determined, and then one of these orders is randomly selected for each participant. This is not as powerful a technique as complete counterbalancing or partial counterbalancing using a Latin squares design. Use of random counterbalancing will result in more random error, but if order effects are likely to be small and the number of conditions is large, this is an option available to researchers.

There are two ways to think about what counterbalancing accomplishes. One is that it controls the order of conditions so that it is no longer a confounding variable. Instead of the attractive condition always being first and the unattractive condition always being second, the attractive condition comes first for some participants and second for others. Likewise, the unattractive condition comes first for some participants and second for others. Thus, any overall difference in the dependent variable between the two conditions cannot have been caused by the order of conditions. A second way to think about what counterbalancing accomplishes is that if there are carryover effects, it makes it possible to detect

them. One can analyze the data separately for each order to see whether it had an effect.

When 9 Is "Larger" Than 221

Researcher Michael Birnbaum (1999) has argued that the lack of context provided by between-subjects designs is often a bigger problem than the context effects created by within-subjects designs. To demonstrate this problem, he asked participants to rate two numbers on how large they were on a scale of 1-to-10 where 1 was "very very small" and 10 was "very very large."

One group of participants were asked to rate the number 9, and another group was asked to rate the number 221 (Birnbaum, 1999). Participants in this between-subjects design gave the number 9 a mean rating of 5.13 and the number 221 a mean rating of 3.10. In other words, they rated 9 as larger than 221! According to Birnbaum, this difference is because participants spontaneously compared 9 with other one-digit numbers (in which case it is relatively large) and compared 221 with other three-digit numbers (in which case it is relatively small).

Experimentation and Validity

When we read about psychology experiments with a critical view, one question to ask is "Is this study valid (accurate)?" However, that question is not as straightforward as it seems because, in psychology, there are many different kinds of validities. Researchers have focused on four validities to help assess whether an experiment is sound: internal validity, external validity, construct validity, and statistical validity (Judd & Kenny, 1981; Morling, 2014). We will explore each validity in depth.

Internal Validity

Two variables being statistically related does not necessarily mean that one causes the other. We have already come across the phrase, "Correlation does not imply causation." For example, if it were the case that people who exercise regularly are happier than people who do not exercise regularly, this implication would not necessarily mean that exercising increases people's happiness. Instead, it could mean that greater happiness causes people to exercise or that something like better physical health causes people to exercise and be happier.

The purpose of an experiment, however, is to show that two variables are statistically related and to do so in a way that supports the conclusion that the independent variable caused any observed differences in the dependent variable. The logic is based on this assumption: If the researcher creates two or more highly similar conditions and then manipulates the independent variable to produce just one difference between them, then any later difference between the conditions must have been caused by the independent variable.

An empirical study is said to be high in internal validity if the

way it was conducted supports the conclusion that the independent variable caused any observed differences in the dependent variable. Thus experiments are high in internal validity because the way they are conducted—with the manipulation of the independent variable and the control of extraneous variables (such as through the use of random assignment to minimize confounds)—provides strong support for causal conclusions. In contrast, non-experimental research designs (e.g., correlational designs), in which variables are measured but are not manipulated by an experimenter, are low in internal validity.

External Validity

At the same time, the way that experiments are conducted sometimes leads to a different kind of criticism. Specifically, the need to manipulate the independent variable and control extraneous variables means that experiments are often conducted under conditions that seem artificial (Bauman et al., 2014). In many psychology experiments, the participants are all undergraduate students and come to a classroom or laboratory to fill out a series of paper-and-pencil questionnaires or perform a carefully designed computerized task. Consider, for example, an experiment in which researcher Barbara Fredrickson and her colleagues (1998) had undergraduate students come to a laboratory on campus and complete a math test while wearing a swimsuit. At first, this manipulation might seem silly. When will undergraduate students ever have to complete math tests in their swimsuits outside of this experiment?

The issue we are confronting is that of external validity. An empirical study is high in external validity if the way it was conducted supports generalizing the results to people and situations beyond those actually studied. As a general rule, studies are higher in external validity when the participants and the

situation studied are similar to those that the researchers want to generalize to and participants encounter every day, often described as mundane realism.

For example, imagine that a group of researchers is interested in how shoppers in large grocery stores are affected by whether breakfast cereal is packaged in yellow or purple boxes. Their study would be high in external validity and have high mundane realism if they studied the decisions of ordinary people doing their weekly shopping in a real grocery store. If the shoppers bought much more cereal in purple boxes, the researchers would be fairly confident that this increase would be true for other shoppers in other stores. Their study would be relatively low in external validity, however, if they studied a sample of undergraduate students in a laboratory at a selective university who merely judged the appeal of various colours presented on a computer screen. On the other hand, this study would have high psychological realism, where the same mental process is used in both the laboratory and in the real world. If the students judged purple to be more appealing than yellow, the researchers would not be very confident that this preference is relevant to grocery shoppers' cereal-buying decisions because of low external validity, but they could be confident that the visual processing of colours has high psychological realism.

We should be careful, however, not to draw the blanket conclusion that experiments are low in external validity. One reason is that experiments need not seem artificial. Consider that Darley and Latané's (1968) experiment provided a reasonably good simulation of a real emergency situation. Or consider field experiments that are conducted entirely outside the laboratory.

In one such experiment, Robert Cialdini (2005) and his colleagues studied whether hotel guests choose to reuse their towels for a second day as opposed to having them washed as a way of conserving water and energy. These researchers manipulated the message on a card left in a large sample of hotel rooms. One version of the message emphasized showing respect for the environment, another emphasized that the hotel would donate a portion of their

savings to an environmental cause, and a third emphasized that most hotel guests choose to reuse their towels. The result was that guests who received the message that most hotel guests choose to reuse their towels, reused their own towels substantially more often than guests receiving either of the other two messages. Given the way they conducted their study, it seems very likely that their result would hold true for other guests in other hotels.

A second reason not to draw the blanket conclusion that experiments are low in external validity is that they are often conducted to learn about psychological processes that are likely to operate in a variety of people and situations. Let us return to the experiment by Fredrickson and colleagues. They found that the women in their study, but not the men, performed worse on the math test when they were wearing swimsuits (Fredrickson et al., 1998). They argued that this gender difference was due to women's greater tendency to objectify themselves—to think about themselves from the perspective of an outside observer—which diverts their attention away from other tasks. They argued, furthermore, that this process of self-objectification and its effect on attention is likely to operate in a variety of women and situations—even if none of them ever finds herself taking a math test in her swimsuit.

Construct Validity

In addition to the generalizability of the results of an experiment, another element to scrutinize in a study is the quality of the experiment's manipulations or the construct validity. The research question that Darley and Latané (1968) started with is "Does helping behavior become diffused?" They hypothesized that participants in a lab would be less likely to help when they believed there were more potential helpers besides themselves. This conversion from research question to experiment design is called operationalization

(see Chapter 4 for more information about the operational definition). Darley and Latané operationalized the independent variable of diffusion of responsibility by increasing the number of potential helpers. In evaluating this design, we would say that the construct validity was very high because the experiment's manipulations very clearly speak to the research question; there was a crisis, a way for the participant to help, and increasing the number of other students involved in the discussion, so they provided a way to test diffusion.

What if the number of conditions in Darley and Latané's study changed? Consider if there were only two conditions: one student involved in the discussion or two. Even though we may see a decrease in helping by adding another person, it may not be a clear demonstration of diffusion of responsibility, just merely the presence of others. We might think it was a form of Bandura's concept of social inhibition. The construct validity would be lower. However, had there been five conditions, perhaps we would see the decrease continue with more people in the discussion or perhaps it would plateau after a certain number of people. In that situation, we may develop a more nuanced understanding of the phenomenon. But by adding more conditions, the construct validity may not get higher. When designing your own experiment, consider how well the research question is operationalized your study.

Statistical Validity

Statistical validity concerns the proper statistical treatment of data and the soundness of the researchers' statistical conclusions. There are many different types of inferential statistics tests (e.g., t-tests, ANOVA, regression, correlation) and statistical validity concerns the use of the proper type of test to analyze the data. When considering the proper type of test, researchers must consider the scale of measure their dependent variable was measured on and the design

of their study. Further, many inferential statistics tests carry certain assumptions (e.g., the data are normally distributed) and statistical validity is threatened when these assumptions are not met; however, the statistics are used nonetheless.

One common critique of experiments is that a study did not have enough participants. The main reason for this criticism is that it is difficult to generalize about a population from a small sample. At the outset, it seems as though this critique is about external validity, but there are studies where small sample sizes are not a problem (subsequent chapters will discuss how small samples, even of only one person, are still very illuminating for psychological research). Therefore, small sample sizes are actually a critique of statistical validity. The statistical validity speaks to whether the statistics conducted in the study are sound and support the conclusions that are made.

The proper statistical analysis should be conducted on the data to determine whether the difference or relationship that was predicted was indeed found. Interestingly, the likelihood of detecting an effect of the independent variable on the dependent variable depends not just on whether a relationship really exists between these variables but also on the number of conditions and the size of the sample. This is why it is important to conduct a power analysis when designing a study, which is a calculation that informs you of the number of participants you need to recruit to detect an effect of a specific size.

Prioritizing Validities

These four big validities–internal, external, construct, and statistical–are useful to keep in mind when both reading about other experiments and designing your own. However, researchers must prioritize, and often it is not possible to have high validity in all four areas. In Cialdini's (2008) study on towel usage in hotels,

the external validity was high but the statistical validity was more modest. This discrepancy does not invalidate the study but it shows where there may be room for improvement for future follow-up studies. Morling (2014) points out that many psychology studies have high internal and construct validity but sometimes sacrifice external validity.

Practical Considerations

The information presented so far in this chapter is enough to design a basic experiment. When it comes time to conduct that experiment, however, several additional practical issues arise. In this section, we consider some of these issues and how to deal with them. Much of this information applies to non-experimental studies as well as experimental ones.

Recruiting Participants

Of course, at the start of any research project, you should be thinking about how you will obtain your participants. Unless you have access to people with schizophrenia or incarcerated juvenile offenders, for example, then there is no point designing a study that focuses on these populations. But even if you plan to use a convenience sample, you will have to recruit participants for your study.

There are several approaches to recruiting participants. One is to use participants from a formal subject pool—an established group of people who have agreed to be contacted about participating in research studies. For example, at many colleges and universities, there is a subject pool consisting of students enrolled in introductory psychology courses who must participate in a certain number of studies to meet a course requirement. Researchers post descriptions of their studies and students sign up to participate, usually via an online system.

Participants who are not in subject pools can also be recruited by posting or publishing advertisements or making personal appeals to groups that represent the population of interest. For example, a researcher interested in studying older adults could arrange to

speak at a meeting of the residents at a retirement community to explain the study and ask for volunteers.

Figure 8.3 *Study (xkcd.com)*

The Volunteer Subject

Even if the participants in a study receive compensation in the form of course credit, a small amount of money, or a chance at being treated for a psychological problem, they are still essentially volunteers. This is worth considering because people who volunteer to participate in psychological research have been shown to differ in predictable ways from those who do not volunteer. Specifically, there is good evidence that on average, volunteers have the following characteristics compared with non-volunteers (Rosenthal & Rosnow, 1976):

- They are more interested in the topic of the research.
- They are more educated.
- They have a greater need for approval.
- They have higher IQ.
- They are more sociable.
- They are higher in social class.

This difference can be an issue of external validity if there is a reason to believe that participants with these characteristics are likely to behave differently than the general population. For example, in testing different methods of persuading people, a rational argument might work better on volunteers than it does on the general population because of their generally higher educational level and IQ.

In many field experiments, the task is not recruiting participants but selecting them. For example, researchers Nicolas Guéguen and Marie-Agnès de Gail (2003) conducted a field experiment on the effect of being smiled at on helping, in which the participants were shoppers at a supermarket. A confederate walking down a stairway

gazed directly at a shopper walking up the stairway and either smiled or did not smile. Shortly afterward, the shopper encountered another confederate, who dropped some computer diskettes on the ground. The dependent variable was whether or not the shopper stopped to help pick up the diskettes.

There are two aspects of this study that are worth addressing here. First, notice that these participants were not "recruited," which means that the institutional review board would have taken care to ensure that dispensing with informed consent in this case was acceptable (e.g., the situation would not have been expected to cause any harm and the study was conducted in the context of people's ordinary activities) (Guéguen & de Gail, 2003).

Second, even though informed consent was not necessary, the researchers still had to select participants from among all the shoppers taking the stairs that day. It is extremely important that this kind of selection be done according to a well-defined set of rules that are established before the data collection begins and can be explained clearly afterward. In this case, with each trip down the stairs, the confederate was instructed to gaze at the first person he encountered who appeared to be between the ages of 20 and 50. Only if the person gazed back did they become a participant in the study.

The point of having a well-defined selection rule is to avoid bias in the selection of participants. For example, if the confederate was free to choose which shoppers he would gaze at, he might choose friendly-looking shoppers when he was set to smile and unfriendly-looking ones when he was not set to smile. As we will see shortly, such biases can be entirely unintentional.

Standardizing the Procedure

It is surprisingly easy to introduce extraneous variables during the procedure. For example, the same experimenter might give clear

instructions to one participant but vague instructions to another. Or one experimenter might greet participants warmly while another barely makes eye contact with them. To the extent that such variables affect participants' behaviour, they add noise to the data and make the effect of the independent variable more difficult to detect.

If they vary systematically across conditions, they become confounding variables and provide alternative explanations for the results. For example, if participants in a treatment group are tested by a warm and friendly experimenter and participants in a control group are tested by a cold and unfriendly one, then what appears to be an effect of the treatment might actually be an effect of experimenter demeanor. When there are multiple experimenters, the possibility of introducing extraneous variables is even greater but is often necessary for practical reasons.

Experimenter's Sex as an Extraneous Variable

It is well known that whether research participants are male or female can affect the results of a study. But what about whether the experimenter is male or female? There is plenty of evidence that this matters too. Male and female experimenters have slightly different ways of interacting with their participants, and of course, participants also respond differently to male and female experimenters (Rosenthal, 1976).

For example, in a recent study on pain perception, participants immersed their hands in icy water for as long as they could (Ibolya et al., 2004). Male participants tolerated the pain longer when the experimenter was a

woman, and female participants tolerated it longer when the experimenter was a man.

Researcher Robert Rosenthal has spent much of his career showing that this kind of unintended variation in the procedure does, in fact, affect participants' behaviour. Furthermore, one important source of such variation is the experimenter's expectations about how participants "should" behave in the experiment. This outcome is referred to as an experimenter expectancy effect (Rosenthal, 1976). For example, if an experimenter expects participants in a treatment group to perform better on a task than participants in a control group, then they might unintentionally give the treatment group participants clearer instructions, more encouragement, or allow them more time to complete the task.

In a striking example, Rosenthal and Kermit Fode (1963) had several students in a laboratory course in psychology train rats to run through a maze. Although the rats were genetically similar, some of the students were told that they were working with "maze-bright" rats that had been bred to be good learners, and other students were told that they were working with "maze-dull" rats that had been bred to be poor learners. Sure enough, over five days of training, the "maze-bright" rats made more correct responses, made the correct response more quickly, and improved more steadily than the "maze-dull" rats.

Clearly, it had to have been the students' expectations about how the rats would perform that made the difference. But how? Some clues come from data gathered at the end of the study, which showed that students who expected their rats to learn quickly felt more positively about their animals and reported behaving toward them in a more friendly manner (e.g., handling them more) (Rosenthal & Fode, 1963).

The way to minimize unintended variation in the procedure is to standardize it as much as possible so that it is carried out in the

same way for all participants regardless of the condition they are in. Here are several ways to do this:

- Create a written protocol that specifies everything that the experimenters are to do and say from the time they greet participants to the time they dismiss them.
- Create standard instructions that participants read themselves or that are read to them word for word by the experimenter.
- Automate the rest of the procedure as much as possible by using software packages for this purpose or even simple computer slide shows.
- Anticipate participants' questions and either raise and answer them in the instructions or develop standard answers for them.
- Train multiple experimenters on the protocol together and have them practice on each other.
- Be sure that each experimenter tests participants in all conditions.

Another good practice is to arrange for the experimenters to be "blind" to the research question or to the condition in which each participant is tested. The idea is to minimize experimenter expectancy effects by minimizing the experimenters' expectations. For example, in a drug study in which each participant receives the drug or a placebo, it is often the case that neither the participants nor the experimenter who interacts with the participants knows which condition they have been assigned to complete. Because both the participants and the experimenters are blind to the condition, this technique is referred to as a double-blind study. (A single-blind study is one in which only the participant is blind to the condition.)

Of course, there are many times this blinding is not possible. For example, if you are both the investigator and the only experimenter, it is not possible for you to remain blind to the research question. Also, in many studies, the experimenter must know the condition

because they must carry out the procedure in a different way in the different conditions.

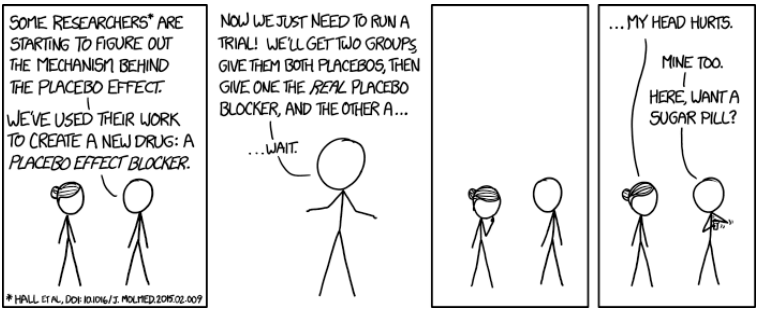


Figure 8.4 *Placebo blocker (xkcd.com)* CC BY-NC 2.5

Record Keeping

It is essential to keep good records when you conduct an experiment. As discussed earlier, it is typical for experimenters to generate a written sequence of conditions before the study begins and then test each new participant in the next condition in the sequence. As you test them, it is a good idea to add to this list the participant's basic demographic information; the date, time, and place of testing; and the name of the experimenter who did the testing. It is also a good idea to have a place for the experimenter to write down comments about unusual occurrences (e.g., a confused or uncooperative participant) or questions that come up. This kind of information can be useful later if you decide to analyze sex differences or effects of different experimenters, or if a question arises about a particular participant or testing session.

Since participants' identities should be kept as confidential (or anonymous) as possible, their names and other identifying information should not be included with their data. In order to identify individual participants, it can, therefore, be useful to assign an identification number to each participant as you test them.

Simply numbering them consecutively beginning with 1 is usually sufficient. This number can then also be written on any response sheets or questionnaires that participants generate, making it easier to keep them together.

Manipulation Check

In many experiments, the independent variable is a construct that can only be manipulated indirectly. For example, a researcher might try to manipulate participants' stress levels indirectly by telling some of them that they have five minutes to prepare a short speech that they will then have to give to an audience of other participants. In such situations, researchers often include a manipulation check in their procedure.

A manipulation check is a separate measure of the construct the researcher is trying to manipulate. The purpose of a manipulation check is to confirm that the independent variable was, in fact, successfully manipulated. For example, researchers trying to manipulate participants' stress levels might give them a paper-and-pencil stress questionnaire or take their blood pressure—perhaps right after the manipulation or at the end of the procedure—to verify that they successfully manipulated this variable.

Manipulation checks are particularly important when the results of an experiment turn out null. In cases where the results show no significant effect of the manipulation of the independent variable on the dependent variable, a manipulation check can help the experimenter determine whether the null result is due to a real absence of an effect of the independent variable on the dependent variable or if it is due to a problem with the manipulation of the independent variable.

Imagine, for example, that you exposed participants to happy or sad movie music—intending to put them in happy or sad moods—but you found that this had no effect on the number of

happy or sad childhood events they recalled. This could be because being in a happy or sad mood has no effect on memories for childhood events. But it could also be that the music was ineffective at putting participants in happy or sad moods. A manipulation check—in this case, a measure of participants' moods—would help resolve this uncertainty. If it showed that you had successfully manipulated participants' moods, then it would appear that there is indeed no effect of mood on memory for childhood events. But if it showed that you did not successfully manipulate participants' moods, then it would appear that you need a more effective manipulation to answer your research question.

Manipulation checks are usually done at the end of the procedure to be sure that the effect of the manipulation lasted throughout the entire procedure and to avoid calling unnecessary attention to the manipulation (to avoid a demand characteristic). However, researchers are wise to include a manipulation check in a pilot test of their experiment so that they avoid spending a lot of time and resources on an experiment that is doomed to fail and instead spend that time and energy on finding a better manipulation of the independent variable.

Pilot Testing

It is always a good idea to conduct a pilot test of your experiment. A pilot test is a small-scale study conducted to make sure that a new procedure works as planned. In a pilot test, you can recruit participants formally (e.g., from an established participant pool) or you can recruit them informally from among family, friends, classmates, and so on. The number of participants can be small, but it should be enough to give you confidence that your procedure works as planned. There are several important questions that you can answer by conducting a pilot test:

- Do participants understand the instructions?
- What kind of misunderstandings do participants have, what kind of mistakes do they make, and what kind of questions do they ask?
- Do participants become bored or frustrated?
- Is an indirect manipulation effective? (You will need to include a manipulation check.)
- Can participants guess the research question or hypothesis (are there demand characteristics)?
- How long does the procedure take?
- Are computer programs or other automated procedures working properly?
- Are data being recorded correctly?

Of course, to answer some of these questions, you will need to observe participants carefully during the procedure and talk with them about it afterward. Participants are often hesitant to criticize a study in front of the researcher, so be sure they understand that their participation is part of a pilot test and you are genuinely interested in feedback that will help you improve the procedure. If the procedure works as planned, then you can proceed with the actual study. If there are problems to be solved, you can solve them, pilot test the new procedure, and continue with this process until you are ready to proceed.

Media Attributions

Summary

Key Takeaways

- An experiment is a type of empirical study that features the manipulation of an independent variable, the measurement of a dependent variable, and control of extraneous variables.
- An extraneous variable is any variable other than the independent and dependent variables. A confound is an extraneous variable that varies systematically with the independent variable.
- Experimental research on the effectiveness of a treatment requires both a treatment condition and a control condition, which can be a no-treatment control condition, a placebo control condition, or a wait-list control condition. Experimental treatments can also be compared with the best available alternative.
- Experiments can be conducted using either between-subjects or within-subjects designs. Deciding which to use in a particular situation requires careful consideration of the pros and cons of each approach.
- Random assignment to conditions in between-subjects experiments or counterbalancing of orders of conditions in within-subjects experiments is a fundamental element of experimental research. The purpose of these techniques is to control extraneous

variables so that they do not become confounding variables.

- Studies are high in internal validity to the extent that the way they are conducted supports the conclusion that the independent variable caused any observed differences in the dependent variable. Experiments are generally high in internal validity because of the manipulation of the independent variable and control of extraneous variables.
- Studies are high in external validity to the extent that the result can be generalized to people and situations beyond those actually studied. Although experiments can seem "artificial"—and low in external validity—it is important to consider whether the psychological processes under study are likely to operate in other people and situations.
- There are several effective methods you can use to recruit research participants for your experiment, including through formal subject pools, advertisements, and personal appeals. Field experiments require well-defined participant selection procedures.
- It is important to standardize experimental procedures to minimize extraneous variables, including experimenter expectancy effects.
- It is important to conduct one or more small-scale pilot tests of an experiment to be sure that the procedure works as planned.

Jessica/Feb 20: for this section to follow the same format as previous chapters there needs to be an Acknowledgments section which is missing here

References

Bauman, C.W., McGraw, A.P., Bartels, D.M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9), 536-554. https://doi.org/10.1111/spc3.12131

Birnbaum, M.H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods*, 4(3), 243-249. https://doi.org/10.1037/1082-989X.4.3.243

Cialdini, R. (2005, April 24). Don't throw in the towel: Use social influence research. *APS Observer*. http://www.psychologicalscience.org/index.php/publications/observer/2005/april-05/dont-throw-in-the-towel-use-social-influence-research.html

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 4(4, Pt. 1), 377–383. https://doi.org/10.1037/h0025589

Fredrickson, B. L., Roberts, T.-A., Noll, S. M., Quinn, D. M., & Twenge, J. M. (1998). The swimsuit becomes you: Sex differences in self-objectification, restrained eating, and math performance. *Journal of Personality and Social Psychology*, 75(1), 269–284. https://doi.org/10.1037/0022-3514.75.1.269

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3), 472–482. https://doi.org/10.1086/586910

Guéguen, N., & de Gail, M-A. (2003). The effect of smiling on helping behavior: Smiling and good Samaritan behavior. *Communication Reports*, 16(2), 133–140. https://doi.org/10.1080/08934210309384496

Ibolya, K., Brake, A., & Voss, U. (2004). The effect of experimenter

characteristics on pain reports in women and men. *Pain*, 112(1), 142–147. https://doi.org/10.1016/j.pain.2004.08.008

Judd, C. M. & Kenny, D. A. (1981). *Estimating the effects of social interventions*. Cambridge University Press.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E-B., & Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain*, 123(12), 2512–2518. https://doi.org/10.1093/brain/123.12.2512

Manning, R., Levine, M., & Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of the 38 witnesses. *American Psychologist*, 62(6), 555–562. https://doi.org/10.1037/0003-066X.62.6.555

Morling, B. (2014, March 31). Teach your students to become better research consumers. *APS Observer*. http://www.psychologicalscience.org/index.php/publications/observer/2014/april-14/teach-your-students-to-be-better-consumers.html

Moseley, J. B., O'Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., Hollingsworth, J. C., Ashton, C. M., & Wray, N. P. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *The New England Journal of Medicine*, 347(2), 81–88. https://doi.org/10.1056/nejmoa013259

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: Recent advances and current thought. *Annual Review of Psychology*, 59, 565–590. https://doi.org/10.1146/annurev.psych.59.113006.095941

Rosenthal, R. (1976). *Experimenter effects in behavioral research* (Enlarged ed.). Irvington Publishers.

Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on performance of the albino rat. *Behavioral Science*, 8(3), 183–189. https://doi.org/10.1002/bs.3830080302

Rosenthal, R., & Rosnow, R. L. (1976). *The volunteer subject*. Wiley.

Shapiro, A. K., & Shapiro, E. (1999). *The powerful placebo: From ancient priest to modern physician*. Johns Hopkins University Press.

Urbaniak, G. C., & Plous, S. (n.d.). Research randomizer. Retrieved June 20, 2025, from https://www.randomizer.org/

IX. INTERPRETING EVIDENCE

Learning Objectives

- Use frequency tables and histograms to display and interpret the distribution of a variable.
- Compute and interpret the mean, median, and mode of a distribution and identify situations in which the mean, median, or mode is the most appropriate measure of central tendency.
- Compute and interpret the range and standard deviation of a distribution.
- Describe differences between groups in terms of their means and standard deviations, and in terms of Cohen's d.
- Describe correlations between quantitative variables in terms of Pearson's r.
- Write out simple descriptive statistics in American Psychological Association (APA) style.
- Interpret and create simple APA-style figures—including bar graphs, line graphs, and scatterplots.
- Interpret and create simple APA-style tables—including tables of group or condition means and correlation matrices.

—*Ecclesiastes 2:13-14 (King James Version)*

וְרָאִיתִי אֲנִי, שֶׁיֵּשׁ יִתְרוֹן לַחָכְמָה מִן-הַסִּכְלוּת–כְּיִתְרוֹן הָאוֹר, מִן-הַחֹשֶׁךְ.

הֶחָכָם עֵינָיו בְּרֹאשׁוֹ, וְהַכְּסִיל בַּחֹשֶׁךְ הוֹלֵךְ; וְיָדַעְתִּי גַם-אֲנִי, שֶׁמִּקְרֶה אֶחָד יִקֶּרֶה אֶת-כֻּלָּם.

Then I saw that wisdom excelleth folly, as far as light excelleth darkness.

The wise man's eyes *are* in his head; but the fool walketh in darkness: and I myself perceived also that one event happeneth to them all.

In science, unlike most other human endeavours, it does not really matter which university you went to or how many degrees you have when validating any conclusion you might put forward. Evidence is the only criteria for determining whether a conclusion is correct or not. Therefore, we need to consider the basics of data analysis in more detail. We will start with descriptive statistics—a set of techniques for summarizing and displaying the data from a sample. We first look at some of the most common techniques for describing single variables, followed by some of the most common techniques for describing statistical relationships between variables. We then look at how to present descriptive statistics in writing and in the form of tables and graphs that would be appropriate for an American Psychological Association (APA)-style research report.

Describing Single Variables

Research invariably comes down to data. When scientists report on the efficacy of a drug, it is based on data. When journalists talk about poll numbers or the state of the economy, it is data. As the world is complicated, we usually have large sets of data that needs to be summarised in an understandable manner. This is where descriptive and inferential statistics come into play.

Descriptive statistics refers to a set of techniques for summarizing and displaying data. Let us assume here that the data are quantitative and consist of scores on one or more variables for each of several study participants. Although in most cases the primary research question will be about one or more statistical relationships between variables, it is also important to describe each variable individually. For this reason, we begin by looking at some of the most common techniques for describing single variables.

The Distribution of a Variable

Every variable has a **distribution**, which is the way the scores are distributed across the levels of that variable. For example, in a sample of 100 university students, the distribution of the variable "number of siblings" might be such that 10 of them have no siblings, 30 have one sibling, 40 have two siblings, and so on. In the same sample, the distribution of the variable "sex" might be such that 44 have a score of "male" and 56 have a score of "female."

Frequency Tables

One way to display the distribution of a variable is in a **frequency table**. Table 7.1, for example, is a frequency table showing a hypothetical distribution of scores on the Rosenberg Self-Esteem Scale for a sample of 40 college students. The first column lists the values of the variable—the possible scores on the Rosenberg scale—and the second column lists the frequency of each score. This table shows that there were three students who had self-esteem scores of 24, five who had self-esteem scores of 23, and so on. From a frequency table like this, one can quickly see several important aspects of a distribution, including the range of scores (from 15 to 24), the most and least common scores (22 and 17, respectively), and any extreme scores that stand out from the rest.

Table 9.1 Frequency Table Showing a Hypothetical Distribution of Scores on the Rosenberg Self-Esteem Scale

Self-esteem	Frequency
24	3
23	5
22	10
21	8
20	5
19	3
18	3
17	0
16	2
15	1

There are a few other points worth noting about frequency tables.

First, the levels listed in the first column usually go from the

highest at the top to the lowest at the bottom, and they usually do not extend beyond the highest and lowest scores in the data. For example, although scores on the Rosenberg scale can vary from a high of 30 to a low of 0, Table 9.1 only includes levels from 24 to 15 because that range includes all the scores in this particular dataset.

Second, when there are many different scores across a wide range of values, it is often better to create a grouped frequency table, in which the first column lists ranges of values and the second column lists the frequency of scores in each range. Table 9.2, for example, is a grouped frequency table showing a hypothetical distribution of simple reaction times for a sample of 20 participants. In a grouped frequency table, the ranges must all be of equal width, and there are usually between five and 15 of them.

Finally, frequency tables can also be used for categorical variables, in which case the levels are category labels. The order of the category labels is somewhat arbitrary, but they are often listed from the most frequent at the top to the least frequent at the bottom.

Table 9.2 Grouped Frequency Table Showing a Hypothetical Distribution of Simple Reaction Times

Reaction time (ms)	Frequency
241–260	1
221–240	2
201–220	2
181–200	9
161–180	4
141–160	2

Histograms

A **histogram** is a graphical display of a distribution. It presents the

same information as a frequency table but in a way that is even quicker and easier to grasp. The histogram in Figure 9.1 presents the distribution of self-esteem scores in Table 9.1. The x-axis of the histogram represents the variable and the y-axis represents frequency. Above each level of the variable on the x-axis is a vertical bar that represents the number of individuals with that score. When the variable is quantitative, as in this example, there is usually no gap between the bars. When the variable is categorical, however, there is usually a small gap between them. (The gap at 17 in this histogram reflects the fact that there were no scores of 17 in this data set.)

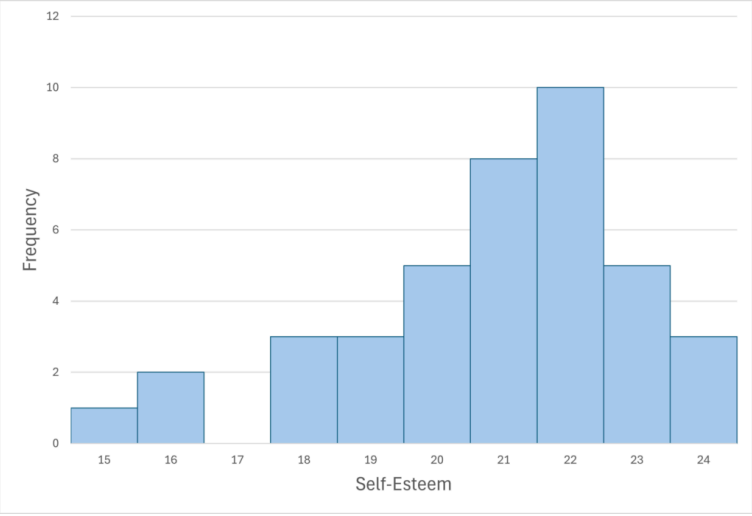


Figure 9.1 *Histogram Showing the Distribution of Self-Esteem Scores Presented in Table 7.1 (Jhangiani et al., 2019)* *Jessica/Feb 20: Image license missing*

Distribution Shapes

When the distribution of a quantitative variable is displayed in a histogram, it has a shape. The shape of the distribution of self-

esteem scores in Figure 9.1 is typical. There is a peak somewhere near the middle of the distribution and "tails" that taper in either direction from the peak. The distribution of Figure 9.1 is unimodal, meaning it has one distinct peak, but distributions can also be bimodal, meaning they have two distinct peaks. Figure 9.2, for example, shows a hypothetical bimodal distribution of scores on the Beck Depression Inventory. Distributions can also have more than two distinct peaks, but these are relatively rare in psychological research.

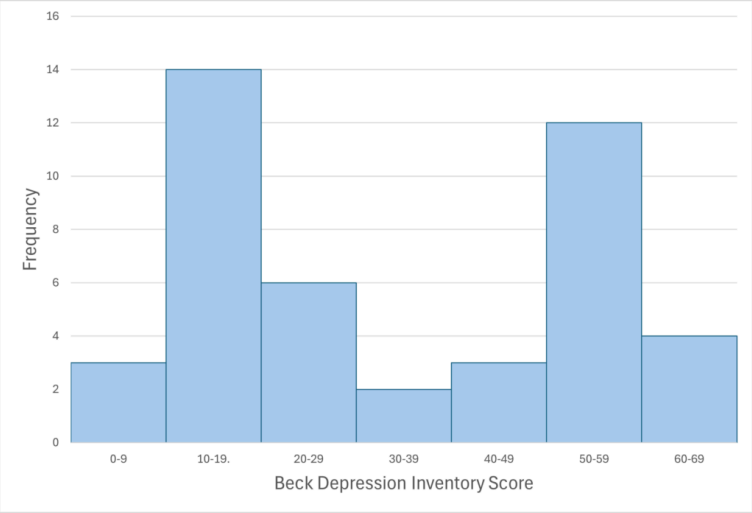


Figure 9.2 *Histogram Showing a Hypothetical Bimodal Distribution of Scores on the Beck Depression Inventory. (Jhangiani et al., 2019)Jessica/Feb 20: Image license missing*

Another characteristic of the shape of a distribution is whether it is symmetrical or skewed. The distribution in the centre of Figure 9.3 is **symmetrical**. Its left and right halves are mirror images of each other. The distribution on the left is negatively **skewed**, with its peak shifted toward the upper end of its range and a relatively long negative tail. The distribution on the right is positively skewed,

with its peak toward the lower end of its range and a relatively long positive tail.

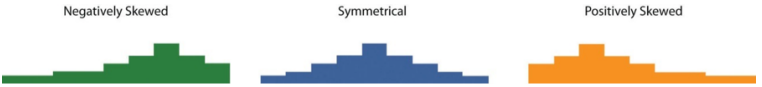


Figure 9.3 *Histograms Showing Negatively Skewed, Symmetrical, and Positively Skewed Distributions (Jhangiani et al., 2019)Jessica/Feb 20: Image license missing*

An **outlier** is an extreme score that is much higher or lower than the rest of the scores in the distribution. Sometimes outliers represent truly extreme scores on the variable of interest. For example, on the Beck Depression Inventory, a single clinically depressed person might be an outlier in a sample of otherwise happy and high-functioning peers. However, outliers can also represent errors or misunderstandings on the part of the researcher or participant, equipment malfunctions, or similar problems.

Media Attributions

- **Figure 7.1** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ
- **Figure 7.2** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ
- **Figure 7.3** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

Measures of Central Tendency and Variability

It is also useful to be able to describe the characteristics of a distribution more precisely. Here, we look at how to do this in terms of two important characteristics: central tendency and variability.

Central Tendency

The **central tendency** of a distribution is its middle—the point around which the scores in the distribution tend to cluster. (Another term for central tendency is *average*.) Looking back at Figure 9.1, for example, we can see that the self-esteem scores tend to cluster around the values 20 to 22. Here, we will consider the three most common measures of central tendency: mean, median, and mode.

The **mean** of a distribution (symbolized M) is the sum of the scores divided by the number of scores. It is an average. As a formula, it looks like this:

\begin{gathered} M=\frac{\Sigma X}{N} \end{gathered}

In this formula, the symbol Σ (the Greek letter sigma) is the summation sign and means to sum across the values of the variable X. N represents the number of scores. The mean is by far the most common measure of central tendency, and there are some good reasons for this. It usually provides a good indication of the central tendency of a distribution, and it is easily understood by most people. In addition, the mean has statistical properties that make it especially useful in doing inferential statistics.

An alternative to the mean is the **median**. The median is the middle score in the sense that half the scores in the distribution are less than it and half are greater than it. The simplest way to find the

median is to organize the scores from lowest to highest and locate the score in the middle. Consider, for example, the following set of seven scores:

8 4 12 14 3 2 3

To find the median, simply rearrange the scores from lowest to highest and locate the one in the middle.

2 3 3 **4** 8 12 14

In this case, the median is 4 because there are three scores lower than 4 and three scores higher than 4. When there is an even number of scores, there are two scores in the middle of the distribution, in which case the median is the value halfway between them. For example, if we were to add a score of 15 to the preceding dataset, there would be two scores (both 4 and 8) in the middle of the distribution, and the median would be halfway between them (6).

One final measure of central tendency is the mode. The **mode** is the most frequent score in a distribution. In the self-esteem distribution presented in Table 9.1 and Figure 9.1, for example, the mode is 22. More students had that score than any other. The mode is the only measure of central tendency that can also be used for categorical variables.

In a distribution that is both unimodal and symmetrical, the mean, median, and mode will be very close to each other at the peak of the distribution. In a bimodal or asymmetrical distribution, the mean, median, and mode can be quite different. In a bimodal distribution, the mean and median will tend to be between the peaks, while the mode will be at the tallest peak.

In a skewed distribution, the mean will differ from the median in the direction of the skew (i.e., the direction of the longer tail). For highly skewed distributions, the mean can be pulled so far in the direction of the skew that it is no longer a good measure of the central tendency of that distribution. Imagine, for example, a set of four simple reaction times of 200, 250, 280, and 250 milliseconds (ms). The mean is 245 ms. But the addition of one more score of 5,000 ms—perhaps because the participant was not paying

attention—would raise the mean to 1,445 ms. Not only is this measure of central tendency greater than 80% of the scores in the distribution, but it also does not seem to represent the behaviour of anyone in the distribution very well. This is why researchers often prefer the median for highly skewed distributions (such as distributions of reaction times).

Keep in mind, though, that you are not required to choose a single measure of central tendency in analyzing your data. Each one provides slightly different information, and all of them can be useful.

Measures of Variability

The **variability** of a distribution is the extent to which the scores vary around their central tendency. Consider the two distributions in Figure 9.4, both of which have the same central tendency. The mean, median, and mode of each distribution are 10. Notice, however, that the two distributions differ in terms of their variability. The top one has relatively low variability, with all the scores relatively close to the centre. The bottom one has relatively high variability, with the scores spread out across a much greater range.

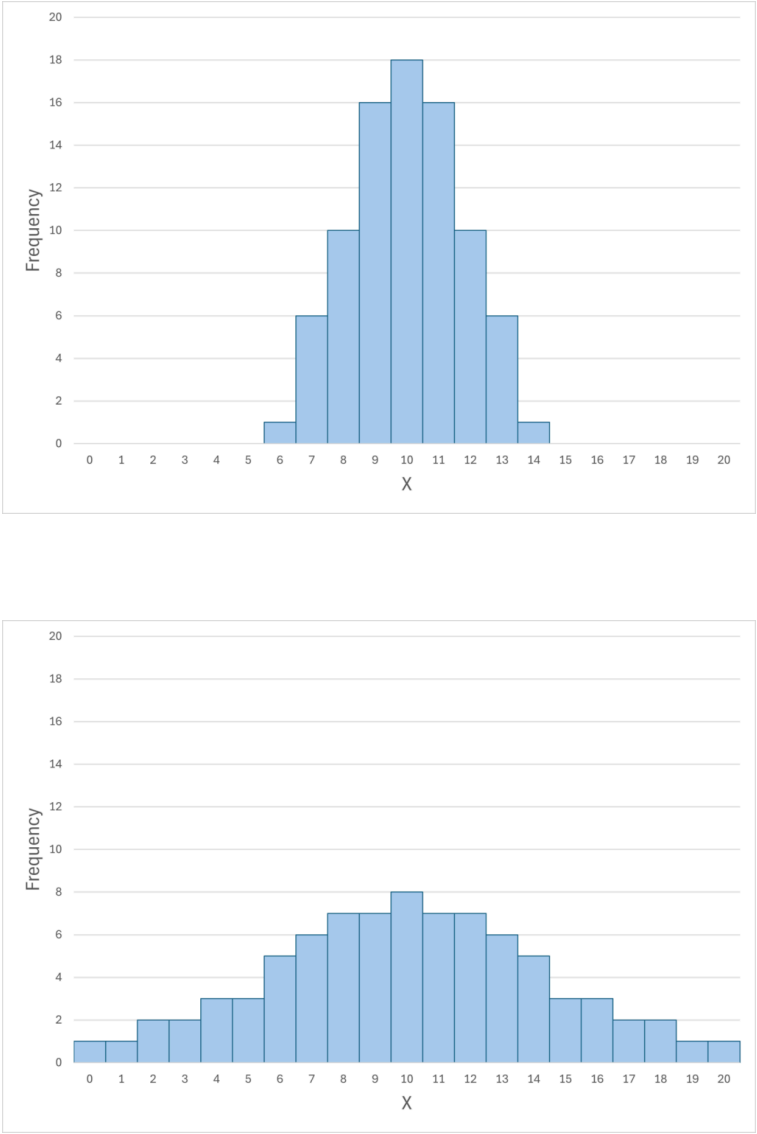


Figure 9.4 *Histograms Showing Hypothetical Distributions With the Same Mean, Median, and Mode (10) but With Low Variability (Top) and High Variability (Bottom). (Jhangiani et al., 2019) Jessica/Feb 20: Image license missing*

One simple measure of variability is the **range**, which is simply the difference between the highest and lowest scores in the distribution. The range of the self-esteem scores in Table 9.1, for example, is the difference between the highest score (24) and the lowest score (15). That is, the range is 24 − 15 = 9. Although the range is easy to compute and understand, it can be misleading when there are outliers. Imagine, for example, an exam in which all the students scored between 90 and 100. It has a range of 10. But if there was a single student who scored 20, the range would increase to 80—giving the impression that the scores were quite variable when, in fact, only one student differed substantially from the rest.

By far the most common measure of variability is the standard deviation. The **standard deviation** of a distribution is the average distance between the scores and the mean. For example, the standard deviations of the distributions in Figure 9.4 are 1.69 for the top distribution and 4.30 for the bottom one. That is, while the scores in the top distribution differ from the mean by about 1.69 units on average, the scores in the bottom distribution differ from the mean by about 4.30 units on average.

Computing the standard deviation involves a slight complication. Specifically, it involves finding the difference between each score and the mean, squaring each difference, finding the mean of these squared differences, and, finally, finding the square root of that mean. We do not need to go into details of how to compute it here; instead, we will discuss what it means and how to interpret it.

Media Attributions

- **Figure 7.4** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

Describing Statistical Relationships

As we have seen throughout this book, most interesting research questions in science are about statistical relationships between variables. In this section, we revisit the two basic forms of statistical relationships introduced earlier in the book—differences between groups or conditions and relationships between quantitative variables—and consider how to describe them in more detail.

Differences Between Groups or Conditions

Differences between groups or conditions are usually described in terms of the mean and standard deviation of each group or condition. For example, Ollendick *et al.* (2009) conducted a study in which they evaluated two one-session treatments for simple phobias. They randomly assigned children with an intense fear (e.g., to dogs) to one of three conditions. In the exposure condition, the children actually confronted the object of their fear under the guidance of a trained therapist. In the education condition, they learned about phobias and some strategies for coping with them. In the wait-list control condition, they were waiting to receive a treatment after the study was over. The severity of each child's phobia was then rated on a 1-to-8 scale by a clinician who did not know which treatment the child had received. (This was one of several dependent variables.) The mean fear rating in the education condition was 4.83 with a standard deviation of 1.52, while the mean fear rating in the exposure condition was 3.47 with a standard deviation of 1.77. The mean fear rating in the control condition was 5.56 with a standard deviation of 1.21. In other words, both

treatments worked, but the exposure treatment worked better than the education treatment.

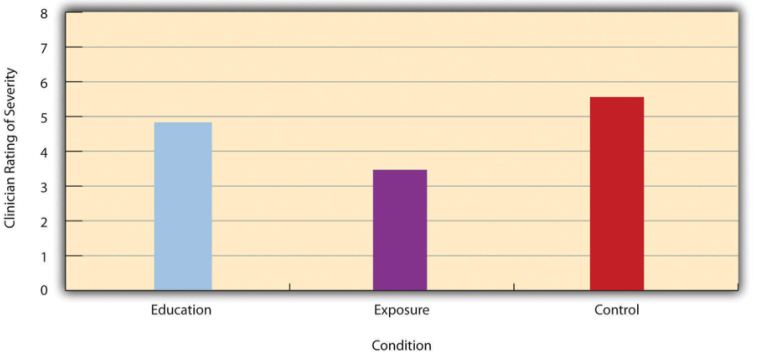


Figure 9.5 *Bar Graph Showing Mean Clinician Phobia Ratings for Children in Two Treatment Conditions. (Jhangiani et al., 2019)Jessica/Feb 20: Image license missing*

It is also important to be able to describe the strength of a statistical relationship, which is often referred to as the **effect size**. The most widely used measure of effect size for differences between group or condition means is called **Cohen's *d***, which is the difference between the two means divided by the standard deviation:

\begin{gathered} d=\frac{(M1-M2)}{SD} \end{gathered}

In this formula, it does not really matter which mean is M1 and which is M2. If there is a treatment group and a control group, the treatment group mean is usually M1 and the control group mean is M2. Otherwise, the larger mean is usually M1 and the smaller mean M2 so that Cohen's *d* turns out to be positive. Indeed, Cohen's *d* values should always be positive so that it is the absolute difference between the means considered in the numerator. The standard deviation in this formula is usually a kind of average of the two groups' standard deviations called the pooled-within group standard deviation. To compute the pooled-within group standard deviation, add the sum of the squared differences for Group 1 to the sum of squared differences for Group 2, divide this by the sum of the

two sample sizes, and then, take the square root of that. Informally, however, the standard deviation of either group can be used instead.

Conceptually, Cohen's *d* is the difference between the two means expressed in standard deviation units. (Notice its similarity to a *z* score, which expresses the difference between an individual score and a mean in standard deviation units.) A Cohen's *d* of 0.50 means that the two groups' means differ by 0.50 standard deviations (half a standard deviation). A Cohen's *d* of 1.20 means that they differ by 1.20 standard deviations.

But how should we interpret these values in terms of the strength of the relationship or the size of the difference between the means? Table 7.4 presents some guidelines for interpreting Cohen's *d* values in psychological research (Cohen, 1992). Values near 0.20 are considered small, values near 0.50 are considered medium, and values near 0.80 are considered large. Thus, a Cohen's *d* value of 0.50 represents a medium-sized difference between two means, and a Cohen's *d* value of 1.20 represents a very large difference in the context of psychological research. In the research by Ollendick *et al.* (2009), there was a large difference (*d* = 0.82) between the exposure and education conditions.

Table 9.4 Guidelines for Referring to Cohen's d and Pearson's r Values as "Strong," "Medium," or "Weak"

Relationship strength	Cohen's *d*	Pearson's *r*
Strong/large	0.80	± 0.50
Medium	0.50	± 0.30
Weak/small	0.20	± 0.10

Cohen's *d* is useful because it has the same meaning regardless of the variable being compared or the scale it was measured on. A Cohen's *d* of 0.20 means that the two group means differ by 0.20 standard deviations whether we are talking about scores on the Rosenberg Self-Esteem Scale, reaction time measured in milliseconds, number of siblings, or diastolic blood pressure

measured in millimetres of mercury. Not only does this make it easier for researchers to communicate with each other about their results, it also makes it possible to combine and compare results across different studies using different measures.

Be aware that the term *effect size* can be misleading because it suggests a causal relationship—that the difference between the two means is an "effect" of being in one group or condition as opposed to another. Imagine, for example, a study showing that a group of exercisers is happier on average than a group of non-exercisers, with an "effect size" of $d = 0.35$. If the study was an experiment—with participants randomly assigned to exercise and no-exercise conditions—then one could conclude that exercising caused a small to medium-sized increase in happiness. If the study was cross-sectional, however, then one could conclude only that the exercisers were happier than the non-exercisers by a small to medium-sized amount. In other words, simply calling the difference an "effect size" does not make the relationship a causal one.

Sex Differences Expressed as Cohen's d

Hyde ***et al.*** (2007) has looked at the results of numerous studies on psychological sex differences and expressed the results in terms of Cohen's *d*. The following are a few of the values she has found, averaging across several studies in each case. (Note that because she always treats the mean for men as *M1* and the mean for women as *M2*, positive values indicate that men score higher and negative values indicate that women score higher.)

Table 9.5 Psychological Sex Differences

Mathematical problem solving	0.08
Reading comprehension	−0.09
Smiling	−0.40
Aggression	0.5
Attitudes toward casual sex	0.81
Leadership effectiveness	−0.02

Hyde points out that although men and women differ by a large amount on some variables (e.g., attitudes toward casual sex), they differ by only a small amount on the vast majority. In many cases, Cohen's *d* is less than 0.10, which she terms a "trivial" difference. (The difference in talkativeness discussed in Chapter 1 was also trivial: $d = 0.06$.) Although researchers and non-researchers alike often emphasize sex differences, Hyde has argued that it makes at least as much sense to think of men and women as fundamentally similar. She refers to this as the "gender similarities hypothesis."

Expressing Your Results

Once you have conducted your descriptive statistical analyses, you will need to present them to others. In this section, we focus on presenting descriptive statistical results in writing, figures, and tables—following American Psychological Association (APA) guidelines for written research reports. These principles can be adapted easily to other presentation formats, such as posters and slideshow presentations.

Media Attributions

- **Figure 7.5** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

Presenting Descriptive Statistics in Writing

Jessica/Feb 20: Images below are missing their license

Recall that APA style includes several rules for presenting numerical results in the text (see 4.31–4.34 in the APA (2020) *Publication Manual*). These include using words only for numbers less than 10 that do not represent precise statistical results and using numerals for numbers 10 and higher. However, statistical results are always presented in the form of numerals rather than words and are usually rounded to two decimal places (e.g., "2.00" rather than "two" or "2"). They can be presented either in the narrative description of the results or parenthetically—much like reference citations. When you have a small number of results to report, it is often most efficient to write them out. Here are some examples:

- The mean age of the participants was 22.43 years with a standard deviation of 2.34.
- Among the participants with low self-esteem, those in a negative mood expressed stronger intentions to have unprotected sex (M= 4.05, SD = 2.32) than those in a positive mood (M = 2.15, SD = 2.27).
- The treatment group had a mean of 23.40 (SD= 9.33), while the control group had a mean of 20.87 (SD = 8.45).
- The test-retest correlation was .96.
- There was a moderate negative correlation between the alphabetical position of respondents' last names and their response time (r= −.27).

Notice that when presented in the narrative, the terms *mean* and *standard deviation* are written out, but when

presented parenthetically, the symbols *M* and *SD* are used instead. Also, notice that it is especially important to use parallel construction to express similar or comparable results in similar ways. The third example is *much* better than the following nonparallel alternative:

- The treatment group had a mean of 23.40 (*SD*= 9.33), while 20.87 was the mean of the control group, which had a standard deviation of 8.45.

Presenting Descriptive Statistics in Figures

When you have a large number of results to report, you can often do it more clearly and efficiently with a graphical depiction of the data, such as pie charts, bar graphs, or scatterplots. In an APA-style research report, these graphs are presented as **figures**.

When you prepare figures for an APA-style research report, there are some general guidelines that you should keep in mind. First, the figure should always add important information rather than repeat information that already appears in the text or in a table (if a figure presents information more clearly or efficiently, then you should keep the figure and eliminate the text or table.) Second, figures should be as simple as possible. For example, the APA (2020) *Publication Manual* discourages the use of colour unless it is absolutely necessary (although colour can still be an effective element in posters, slideshow presentations, or textbooks.) Third, figures should be interpretable on their own. A reader should be able to understand the basic result based only on the figure and its caption and should not have to refer to the text for an explanation.

There are also several more technical guidelines for presentation of figures that include the following (see the APA (2020) *Publication Manual* section 5.20 through 5.30):

- Layout of graphs
    - In general, scatterplots, bar graphs, and line graphs should be slightly wider than they are tall.
    - The independent variable should be plotted on the x-axis and the dependent variable on the y-axis.
    - Values should increase from left to right on the x-axis and from bottom to top on the y-axis.
    - The x-axis and y-axis should begin with the value zero.
- Axis labels and legends
    - Axis labels should be clear and concise and include the units of measurement if they do not appear in the caption.
    - Axis labels should be parallel to the axis.
    - Legends should appear within the figure.
    - Text should be in the same simple font throughout and no smaller than 8 pt and no larger than 14 pt.
- Captions
    - Captions are titled with the word "Figure," followed by the figure number in the order in which it appears in the text and terminated with a period. This title is italicized.
    - After the title is a brief description of the figure terminated with a period (e.g., "Reaction times of the control versus experimental group.")
    - Following the description, include any information needed to interpret the figure, such as any abbreviations, units of measurement (if not in the axis label), units of error bars, etc.

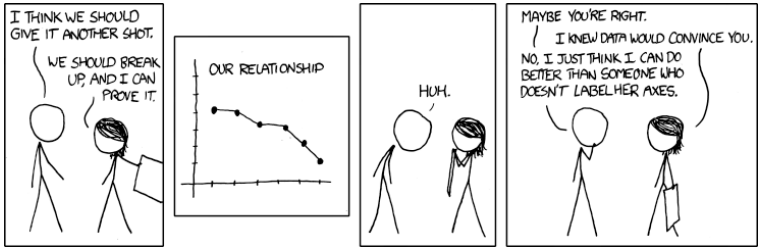


***Figure* 7.6** *retrieved from* XKCD

Bar Graphs

As we have seen throughout this book, **bar graphs** are generally used to present and compare the mean scores for two or more groups or conditions. The bar graph in Figure 9.7 is an APA-style version of Figure 9.7. Notice that it conforms to all the guidelines listed.

A new element in Figure 9.7 is the smaller vertical bars that extend both upward and downward from the top of each main bar. These are **error bars**, and they represent the variability in each group or condition. Although they sometimes extend one standard *deviation* in each direction, they are more likely to extend one standard *error* in each direction (as in Figure 9.7). The **standard error** is the standard deviation of the group divided by the square root of the sample size of the group. The standard error is used because, in general, a difference between group means that is greater than two standard errors is statistically significant. Thus, one can "see" whether a difference is statistically significant based on a bar graph with error bars.

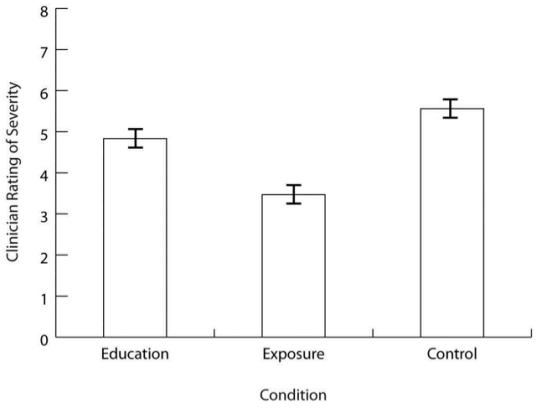


Figure X. Mean clinician's rating of phobia severity for participants receiving the education treatment and the exposure treatment. Error bars represent standard errors.

Figure 9.7 *Sample APA-Style Bar Graph, With Error Bars Representing the Standard Errors Based on Research by Ollendick and Colleagues (from Jhangiani et al., 2019).*

Line Graphs

Line graphs are used when the independent variable is measured in a more continuous manner (e.g., time) or to present correlations between quantitative variables when the independent variable has, or is organized into, a relatively small number of distinct levels. Each point in a line graph represents the mean score on the dependent variable for participants at one level of the independent variable. Figure 7.8 is an APA-style version of the results of Carlson and Conard (2011). Notice that it includes error bars representing the standard error and conforms to all the stated guidelines.

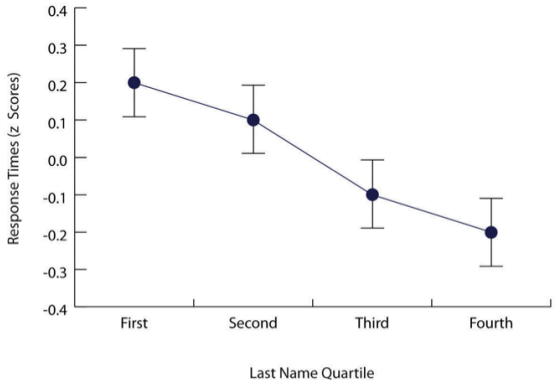


Figure X. Mean response time by the alphabetical position of respondents' names in the alphabet. Response times are expressed as **z** scores. Error bars represent standard errors.

Figure 9.8 *Sample APA-Style Line Graph [Based on Research by Carlson & Conard (2011)]*

In most cases, the information in a line graph could just as easily be presented in a bar graph. In Figure 9.8, for example, one could replace each point with a bar that reaches up to the same level and leave the error bars right where they are. This emphasizes the fundamental similarity of the two types of statistical relationships. Both are differences in the average score of one variable across the levels of another. The convention followed by most researchers, however, is to use a bar graph when the variable plotted on the *x*-axis is categorical and a line graph when it is quantitative.

Scatterplots

Scatterplots are used to present correlations and relationships between quantitative variables when the variable on the *x*-axis (typically the independent variable) has a large number of levels. Each point in a scatterplot represents an individual rather than the

mean for a group of individuals, and there are no lines connecting the points.

The graph in Figure 9.9 illustrates a few points. First, when the variables on the x-axis and y-axis are conceptually similar and measured on the same scale—as here, where they are measures of the same variable on two different occasions—this can be emphasized by making the axes the same length. Second, when two or more individuals fall at exactly the same point on the graph, one way this can be indicated is by offsetting the points slightly along the x-axis. Other ways are by displaying the number of individuals in parentheses next to the point or by making the point larger or darker in proportion to the number of individuals. Finally, the straight line that best fits the points in the scatterplot, which is called the regression line, can also be included.

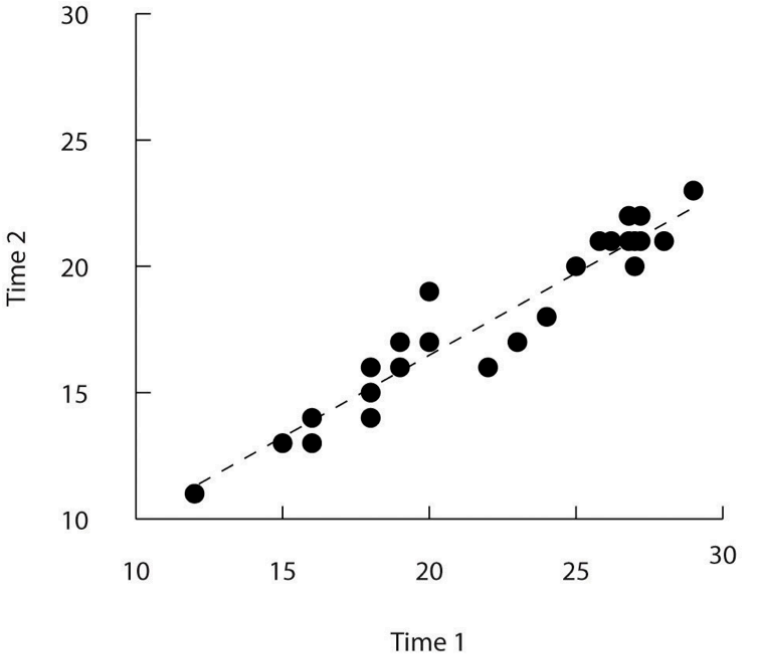


Figure X. Relationship between scores on the Rosenberg self-esteem scale taken by 25 research methods students on two occasions one week apart. Pearson's r =.96.

Figure 9.9 *Sample* APA-Style *Scatterplot. (Jhangiani et al., 2019)*

Expressing Descriptive Statistics in Tables

Like graphs, tables can be used to present large amounts of information clearly and efficiently. The same general principles that apply to tables also apply to graphs. They should add important information to the presentation of your results, be as simple as possible, and be interpretable on their own. Again, this section focuses on tables for an APA-style manuscript specifically.

The most common use of tables is to present several means and

standard deviations—usually for complex research designs with multiple independent and dependent variables. Figure 9.10, for example, shows the results of a hypothetical study similar to the one by MacDonald and Martineau (2002) (The means in Figure 9.10 are the means reported by MacDonald and Martineau, but the standard errors are not). Recall that these researchers categorized participants as having low or high self-esteem, put them into a negative or positive mood, and measured their intentions to have unprotected sex. They also measured participants' attitudes toward unprotected sex.

Notice that the table includes horizontal lines spanning the entire table at the top and bottom, and just beneath the column headings. Furthermore, every column has a heading—including the leftmost column—and there are additional headings that span two or more columns that help to organize the information and present it more efficiently. Finally, notice that APA-style tables are numbered consecutively starting at 1 (Table 1, Table 2, and so on) and given a brief but clear and descriptive title.

Table X

Means and Standard Deviations of Intentions to Have Unprotected Sex and Attitudes Toward Unprotected Sex as a Function of Both Mood and Self-Esteem

	Negative mood		Positive mood	
Self-Esteem	*M*	*SD*	*M*	*SD*
		Intentions		
High	2.46	1.97	2.45	2.00
Low	4.05	2.32	2.15	2.27
		Attitudes		
High	1.65	2.23	1.82	2.32
Low	1.95	2.01	1.23	1.75

Figure 9.10 *Sample APA-Style Table Presenting Means and Standard Deviations.* (*Jhangiani et al.*, 2019)

Another common use of tables is to present correlations—usually measured by Pearson's r—among several variables. This kind of table

is called a **correlation matrix**. Figure 9.11 is a correlation matrix based on a study by David McCabe and colleagues (2010). They were interested in the relationships between working memory and several other variables. We can see from the table that the correlation between working memory and executive function, for example, was an extremely strong .96, the correlation between working memory and vocabulary was a medium .27, and all the measures except vocabulary tend to decline with age.

Notice here that only half the table is filled in because the other half would have identical values. For example, the Pearson's *r* value in the upper right corner (working memory and age) would be the same as the one in the lower left corner (age and working memory). The correlation of a variable with itself is always 1.00, so these values are replaced by dashes to make the table easier to read.

Table X

Correlations Between Five Cognitive Variables and Age

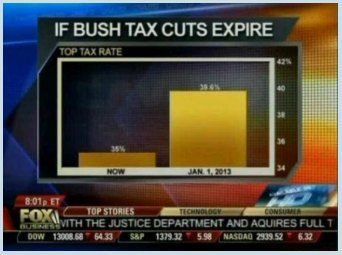
Measure	1	2	3	4	5
1. Working memory	—				
2. Executive function	.96	—			
3. Processing speed	.78	.78	—		
4. Vocabulary	.27	.45	.08	—	
5. Episodic memory	.73	.75	.52	.38	—
6. Age	−.59	−.56	−.82	.22	−.41

Figure 7.11 *Sample APA-Style Table (Correlation Matrix) [Based on Research by McCabe et al. (2010)]*

As with graphs, precise statistical results that appear in a table do not need to be repeated in the text. Instead, the writer can note major trends and alert the reader to details (e.g., specific correlations) that are of particular interest.

Manipulating Graphs

Manipulating graphs can mislead the audience and lead to false conclusions. While this could be done by mistake, this is often done deliberately to manipulate people towards a particular conclusion. Consider some of the following graphs from Fox News:



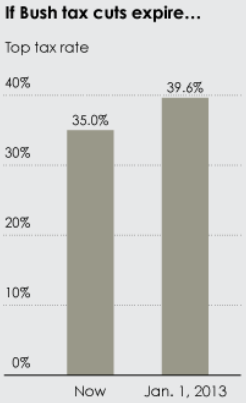


Figure 7.12 Bar graphs showing the effects of expiring Bush tax cuts as presented on Fox News and with proper ratios (Flowing Data, 2012)

In the above graph on the left, we see an actual bar graph broadcast by Fox News on the effects of the expiration of the Bush tax cuts under President Obama. At first glance, it looks like a five-fold increase in tax burden. However, we can see on the graph on the left that if you show the bar graph starting from zero, which shows a modest difference of 4.6%.

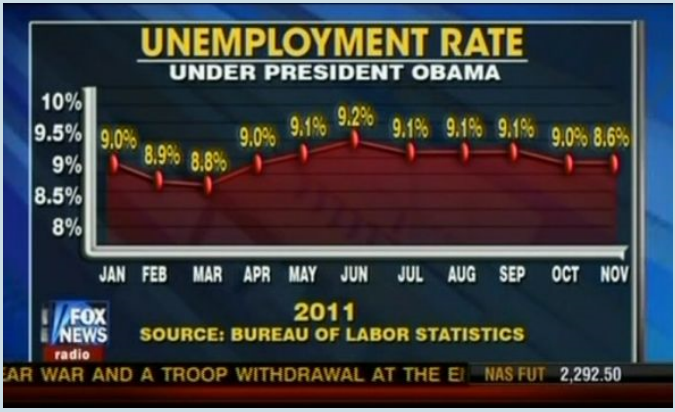


***Figure* 7.13** *Unemployment rate under President Obama presented on Fox News (Flowing Data, 2011)*

The above line graph is supposed to show the unemployment rate under president Obama. However, if we look carefully, we can see that the rises and drops in the graph have no relationship to the actual numbers. A drop from 9% to 8.6% in November is shown as a horizontal line (indicating no change) even though the rate in November is lower than the rate in March. The actual graph should look something like this:

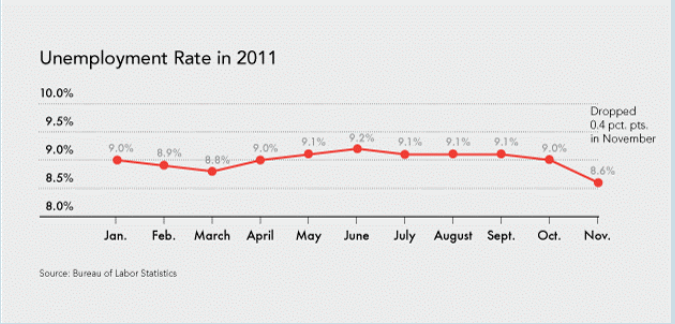


Figure 7.14 *Unemployment rate under President Obama with corrected lines (Flowing Data, 2011)*

These are just a few examples of how partisan media (such as Fox News) can manipulate data to reflect their own biases.

Media Attributions

- **Figure 7.6** http://imgs.xkcd.com/comics/convincing.png (CC-BY-NC 2.5)
- **Figure 7.7** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ
- **Figure 7.8** Carlson, K. A., & Conard, J. M. (2011). The last name effect: How last name influences acquisition timing. *Journal of Consumer Research*, 38(2), 300–307. https://doi.org/10.1086/658470
- **Figure 7.9** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton,

D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ

- **Figure 7.10** Jhangiani, R.S., Chiang, I.A. Cuttler, C. & Leighton, D.C. (2019). Research Methods in Psychology, Surrey, Canada: Kwantlen Polytechnic University. https://doi.org/10.17605/OSF.IO/HF7DQ
- **Figure 7.11** McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. https://doi.org/10.1037/a0017619
- **Figure 7.12** Flowing Data (August, 2012). *Fox News continues charting excellence*. FlowingData. https://flowingdata.com/2012/08/06/fox-news-continues-charting-excellence/
- **Figure 7.13** Flowing Data (December, 2011). *Fox News still makes awesome charts*. FlowingData. https://flowingdata.com/2011/12/12/fox-news-still-makes-awesome-charts/
- **Figure 7.14** Flowing Data (December, 2011). *Fox News still makes awesome charts*. FlowingData. https://flowingdata.com/2011/12/12/fox-news-still-makes-awesome-charts/

Conclusion

Jessica/Feb 20: Title updated from Conclusion to Summary to follow the same format as other chapters. Under textbox we are missing the acknowledgement sections other chapters have

Key Takeaways

- Every variable has a distribution—a way that the scores are distributed across the levels. The distribution can be described using a frequency table and histogram. It can also be described in words in terms of its shape, including whether it is unimodal or bimodal, and whether it is symmetrical or skewed.
- The central tendency, or middle, of a distribution can be described precisely using three statistics—the mean, median, and mode. The mean is the sum of the scores divided by the number of scores, the median is the middle score, and the mode is the most common score.
- The variability, or spread, of a distribution can be described precisely using the range and standard deviation. The range is the difference between the highest and lowest scores, and the standard deviation is the average amount by which the scores differ from the mean.
- Differences between groups or conditions are typically described in terms of the means and standard deviations of the groups or conditions or in

terms of Cohen's *d* and are presented in bar graphs.

- Cohen's *d* is a measure of relationship strength (or effect size) for differences between two group or condition means. It is the difference of the means divided by the standard deviation. In general, values of ±0.20, ±0.50, and ±0.80 can be considered small, medium, and large, respectively.
- Correlations between quantitative variables are typically described in terms of Pearson's *r* and presented in line graphs or scatterplots.
- Pearson's *r* is a measure of relationship strength (or effect size) for relationships between quantitative variables. It is the mean cross-product of the two sets of *z* scores. In general, values of ±.10, ±.30, and ±.50 can be considered small, medium, and large, respectively.
- In an APA-style article, simple results are most efficiently presented in the text, while more complex results are most efficiently presented in graphs or tables.
- APA style includes several rules for presenting numerical results in the text. These include using words only for numbers less than 10 that do not represent precise statistical results, and rounding results to two decimal places, using words (e.g., "mean") in the text and symbols (e.g., "M") in parentheses.
- APA style includes several rules for presenting results in graphs and tables. Graphs and tables should add information rather than repeating information, be as simple as possible, and be interpretable on their own with a descriptive caption (for graphs) or a

descriptive title (for tables).

References

American Psychological Association. (2020). *Publication manual of the American psychological association* (7th ed.).

Carlson, K. A., & Conard, J. M. (2011). The last name effect: How last name influences acquisition timing. *Journal of Consumer Research*, 38(2), 300-307. doi: 10.1086/658470

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.

Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science*, 16, 259–263.

MacDonald, T. K., & Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self-esteem lead to risky health behaviors? *Journal of Experimental Social Psychology*, 38(3), 299–306. https://doi.org/10.1006/jesp.2001.1505

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222–243. https://doi.org/10.1037/a0017619

Ollendick, T. H., Öst, L.-G., Reuterskiöld, L., Costa, N., Cederlund, R., Sirbu, C.,...Jarrett, M. A. (2009). One-session treatments of specific phobias in youth: A randomized clinical trial in the United States and Sweden. *Journal of Consulting and Clinical Psychology*, 77, 504–516.

X. PEER REVIEW

Learning Objectives

- Define “replication.”
- Explain the difference between exact and conceptual replication.
- List explanations for non-replication.
- Name potential solutions to the replication crisis.

Reporting Research

…nam quam aram sibi parare potest, qui rationis majestatem laedit?

What altar of refuge can a man find for himself when he commits treason against the majesty of reason?

— Baruch Spinoza
Theological Political Treatise XV

When scientists complete a research project, they generally want to share their findings with other scientists. The American Psychological Association (APA) publishes a manual detailing how to write a paper for submission to scientific journals. Unlike an article that might be published in a magazine like *Psychology Today*, which targets a general audience with an interest in a subject such as psychology, scientific journals generally publish **peer-reviewed journal articles** aimed at an audience of professionals and scholars who are actively involved in research themselves.

A peer-reviewed journal article is read by several other scientists (generally anonymously) with expertise in the subject matter. These peer reviewers provide feedback—to both the author and the journal editor—regarding the quality of the draft.

While reading, peer reviewers:

- Look for a strong rationale for the research being described, a clear description of how the research was conducted, and evidence that the research was conducted in an ethical manner.
- Look for flaws in the study's design, methods, and statistical analyses.
- Check that the conclusions drawn by the authors seem reasonable given the observations made during the research.
- Comment on how valuable the research is in advancing the discipline's knowledge, which helps prevent unnecessary duplication of research findings in the scientific literature and,

to some extent, ensures that each research article provides new information.

Ultimately, the journal editor will compile all of the peer reviewer feedback and determine whether the article will be published in its current state (a rare occurrence), published with revisions, or not accepted for publication.

Peer review provides some degree of quality control for scientific research. Poorly conceived or executed studies can be weeded out, and even well-designed research can be improved by the revisions suggested. Peer review also ensures that the research is described clearly enough to allow other scientists to **replicate** it, meaning they can repeat the experiment using different samples to determine reliability. Sometimes, replications involve additional measures that expand on the original finding. In any case, each replication serves to provide more evidence to support the original research findings.

Successful replications of published research make scientists more apt to adopt those findings, while repeated failures tend to cast doubt on the legitimacy of the original article and lead scientists to look elsewhere. For example, it would be a major advancement in the medical field if a published study indicated that taking a new drug helped individuals achieve better health without changing their behaviour. But if other scientists could not replicate the results, the original study's claims would be questioned.

In recent years, there has been increasing concern about a "replication crisis" that has affected a number of scientific fields, including psychology. Some of the most well-known studies and scientists have produced research that has failed to be replicated by others (as discussed in Shrout & Rodgers, 2018). In fact, even a famous Nobel Prize-winning scientist has recently retracted a published paper because she had difficulty replicating her results (Nobel Prize-winning scientist Frances Arnold retracts paper (BBC, 2020)). These kinds of outcomes have prompted some scientists to begin to work together and more openly, and some would argue that the current "crisis" is actually improving the ways in which

science is conducted and in how its results are shared with others (Aschwanden, 2018).

The Vaccine-Autism Myth and Retraction of Published Studies

Some scientists have claimed that routine childhood vaccines cause some children to develop autism, and, in fact, several peer-reviewed publications published research making these claims. Since the initial reports, large-scale epidemiological research has indicated that vaccinations are not responsible for causing autism and that it is much safer to have your child vaccinated than not. Furthermore, several of the original studies making this claim have since been retracted.

A published piece of work can be rescinded when data is called into question because of falsification, fabrication, or serious research design problems. Once rescinded, the scientific community is informed that there are serious problems with the original publication. Retractions can be initiated by the researcher who led the study, by research collaborators, by the institution that employed the researcher, or by the editorial board of the journal in which the article was originally published. In the vaccine-autism case, the retraction was made because of a significant conflict of interest in which the leading researcher had a financial interest in establishing a link between childhood vaccines and autism (Offit, 2008). Unfortunately, the initial studies received so much media attention that many parents around the world became hesitant to have their children vaccinated.

Continued reliance on such debunked studies has significant consequences. For instance, between January and October of 2019, there were 22 measles outbreaks across the United States and more than a thousand cases of individuals contracting measles (Patel et al., 2019). This is likely due to the anti-vaccination movements that have risen from the debunked research. For more information about how the vaccine/autism story unfolded, as well as the repercussions of this story, take a look at Paul Offit's book, *Autism's False Prophets: Bad Science, Risky Medicine, and the Search for a Cure.*

The Replication Crisis

In science, replication is the process of repeating research to determine the extent to which findings generalize across time and situations. Recently, the science of psychology has come under criticism because a number of research findings cannot be replicated. In this module, we discuss reasons for non-replication and the impact this phenomenon has on the field as well as suggest solutions to this problem.

The Disturbing Problem

If you were driving down the road and saw a pirate standing at an intersection, you might not believe your eyes. But if you continued driving and saw a second, and then a third, you might become more confident in your observations. The more pirates you saw, the less likely the first sighting would be a false positive (you were driving fast and the person was just wearing an unusual hat and billowy shirt) and the more likely it would be the result of a logical reason (there is a pirate-themed conference in town). This somewhat absurd example is a real-life illustration of replication: the repeated findings of the same results.

The replication of findings is one of the defining hallmarks of science. Scientists must be able to replicate the results of studies, or their findings do not become part of the scientific knowledge. Replication protects against false positives (seeing a result that is not really there) and increases confidence that the result actually exists. If you collect satisfaction data among homeless people living in Kolkata, India, for example, it might seem strange that they would report fairly high satisfaction with their food (Biswas-Diener & Diener, 2001). If you find the exact same result at a *different* time and with a *different* sample of homeless people living in Kolkata,

however, you can feel more confident that this result is true (Biswas-Diener & Diener, 2001).

In modern times, the science of psychology is facing a crisis. It turns out that many studies in psychology—including many highly cited studies—cannot be replicated. In an era where news is instantaneous, the failure to replicate research raises important questions about the scientific process in general and psychology specifically. People have the right to know if they can trust research evidence. For our part, psychologists also have a vested interest in ensuring that our methods and findings are as trustworthy as possible.

Psychology is not alone in coming up short on replication. There have been notable failures to replicate findings in other scientific fields, as well. For instance, in 1989, scientists reported that they had produced "cold fusion," achieving nuclear fusion at room temperatures. This could have been an enormous breakthrough in the advancement of clean energy. However, other scientists were unable to replicate the findings. Thus, the potentially important results did not become part of the scientific canon, and a new energy source did not materialize.

In medical science, as well, a number of findings have been found that cannot be replicated—which is of vital concern to all of society. The non-reproducibility of medical findings suggests that some treatments for illness could be ineffective. One example of non-replication has emerged in the study of genetics and diseases: when replications were attempted to determine whether certain gene-disease findings held up, only about 4% of the findings consistently did so.

The non-reproducibility of findings is disturbing because it suggests the possibility that the original research was done sloppily. Even worse is the suspicion that the research may have been falsified. In science, faking results is *the biggest* of sins, the unforgivable sin, and for this reason, the field of psychology has been thrown into an uproar. However, as we will discuss, there are a number of explanations for non-replication, and not all are bad.

What is Replication?

There are different types of replications. First, there is a type called "**exact replication**" (also called "**direct replication**"). In this form, a scientist attempts to exactly recreate the scientific methods used in conditions of an earlier study to determine whether the results come out the same. If, for instance, you wanted to exactly replicate Asch's (1956) classic findings on conformity, you would follow the original methodology: you would use only male participants, use groups of 8, and present the same stimuli (lines of differing lengths) in the same order. The second type of replication is called "**conceptual replication**." This occurs when—instead of an exact replication, which reproduces the methods of the earlier study as closely as possible—a scientist tries to confirm the previous findings using a different set of specific methods that test the same idea. The same hypothesis is tested but using a different set of methods and measures. A conceptual replication of Asch's research might involve both male and female **confederates** purposefully misidentifying types of fruit to investigate conformity—rather than only males misidentifying line lengths.

Both exact and conceptual replications are important because they each tell us something new. Exact replications tell us whether the original findings are true, at least under the exact conditions tested. Conceptual replications help confirm whether the theoretical idea behind the findings is true, and under what conditions these findings will occur. In other words, conceptual replication offers insights into how generalizable the findings are.

Enormity of the Current Crisis

Recently, there has been growing concern as psychological research fails to be replicated. To give you an idea of the extent of non-

replicability of psychology findings, Table 10.1 shows data reported in 2015 by the Open Science Collaboration project, led by University of Virginia psychologist Brian Nosek (Open Science Collaboration, 2015). Because these findings were reported in the prestigious journal, *Science*, they received widespread attention from the media. Here are the percentages of research that was replicated—selected from several highly prestigious journals:

Table 10.1 The Reproducibility of Psychological Science

Journal	**% of Findings Replicated**
Journal of Personality and Social Psychology: Social	23
Journal of Experimental Psychology: Learning, Memory, and Cognition	48
Psychological Science, social articles	29
Psychological Science, cognitive articles	53
Overall	**36**

Clearly, there is a very large problem when only about one-third of the psychological studies in premier journals replicate! It appears that this problem is particularly pronounced for social psychology, but even the 53% replication level of cognitive psychology is cause for concern.

The situation in psychology has grown so worrisome that the Nobel Prize-winning psychologist Daniel Kahneman called on social psychologists to clean up their act (Kahneman, 2012). The Nobel laureate spoke bluntly of doubts about the integrity of psychology research, calling the current situation in the field a "mess." His missive was pointed primarily at researchers who study social "priming," but in light of the non-replication results that have since come out, it might be more aptly directed at the behavioural sciences in general.

Examples of Non-replications in Psychology

A large number of scientists have attempted to replicate studies on what might be called "metaphorical priming," and more often than not, these replications have failed. **Priming** is the process by which a recent reference (often a subtle, subconscious cue) can increase the accessibility of a trait. For example, if your instructor says, "Please put aside your books, take out a clean sheet of paper, and write your name at the top," you might find your pulse quickening. Over time, you have learned that this cue means you are about to be given a pop quiz. This phrase primes all the features associated with pop quizzes: they are anxiety-provoking, they are tricky, and your performance matters.

One example of a priming study that, at least in some cases, does not replicate is the priming of the idea of intelligence. In theory, it might be possible to prime people to actually become more intelligent (or perform better on tests, at least). For instance, in one study, priming students with the idea of a stereotypical professor versus soccer hooligans led participants in the "professor" condition to earn higher scores on a trivia game (Dijksterhuis & van Knippenberg, 1998). Unfortunately, in several follow-up instances, this finding has not been replicated (Shanks et al., 2013). This is unfortunate for all of us because it would be a very easy way to raise our test scores and general intelligence. If only it were true.

Another example of a finding that seems to not replicate consistently is the use of spatial distance cues to prime people's feelings of emotional closeness to their families (Williams & Bargh, 2008). In this type of study, participants are asked to plot points on graph paper, either close together or far apart. The participants are then asked to rate how close they are to their family members. Although the original researchers found that people who plotted close-together points on graph paper reported being closer to their relatives, studies reported on PsychFileDrawer—an internet repository of replication attempts—suggest that the findings

frequently do not replicate. Again, this is unfortunate because it would be a handy way to help people feel closer to their families.

As one can see from the examples, some of the studies that fail to replicate report extremely interesting findings—even counterintuitive findings that appear to offer new insights into the human mind. Critics claim that psychologists have become too enamored with such newsworthy, surprising "discoveries" that receive a lot of media attention. This raises the question of timing: might the current crisis of non-replication be related to the modern, media-hungry context in which psychological research (indeed, all research) is conducted? Put another way: is the non-replication crisis new?

Nobody has tried to systematically replicate studies from the past, so we do not know if published studies are becoming less replicable over time. In 1990, however, Amir and Sharon were able to successfully replicate most of the main effects of six studies from another culture, though they did fail to replicate many of the interactions. This particular shortcoming in their overall replication may suggest that published studies are becoming less replicable over time, but we cannot be certain. What we can be sure of is that there is a significant problem with replication in psychology, and it is a trend the field needs to correct. Without replicable findings, nobody will be able to believe in scientific psychology.

Reasons for Non-replication

When findings do not replicate, the original scientists sometimes become indignant and defensive, offering reasons or excuses for the non-replication of their findings—including, at times, attacking those attempting the replication. They sometimes claim that the scientists attempting the replication are unskilled or unsophisticated, or do not have sufficient experience to replicate

the findings. This, of course, might be true, and it is one possible reason for non-replication.

Although many believe that the failure to replicate research results is an expected characteristic of cumulative scientific progress, others have interpreted this situation as evidence of systematic problems with conventional scholarship in psychology, including a publication bias that favours the discovery and publication of counter-intuitive but statistically significant findings instead of the duller (but incredibly vital) process of replicating previous findings to test their robustness (Aschwanden, 2015; Frank, 2015; Pashler & Harris, 2012; Scherer, 2015). Worse still is the suggestion that the low replicability of many studies is evidence of the widespread use of questionable research practices by psychological researchers.

Questionable research practices include:

1. The selective deletion of outliers in order to influence (usually by artificially inflating) statistical relationships among the measured variables.
2. The selective reporting of results, cherry-picking only those findings that support one's hypotheses.
3. Mining the data without an *a priori* hypothesis, only to claim that a statistically significant result had been originally predicted, a practice referred to as "**HARKing**" or hypothesizing after the results are known (Kerr, 1998).
4. A practice colloquially known as "***p*-hacking**" (briefly discussed in the previous section), in which a researcher might perform inferential statistical calculations to see if a result was significant before deciding whether to recruit additional participants and collect more data (Head et al., 2015). As you have learned, the probability of finding a statistically significant result is influenced by the number of participants in the study.
5. Outright fabrication of data, although this would be a case of fraud rather than a "research practice."

One reason for defensive responses is the unspoken implication that the original results might have been falsified. Faked results are only one reason studies may not replicate, but it is the most disturbing reason. We hope faking is rare, but in the past decade, a number of shocking cases have turned up. Perhaps the most well-known ones come from social psychology. Diederik Stapel, a renowned social psychologist in the Netherlands, admitted to faking the results of a number of studies. Marc Hauser, a popular professor at Harvard, apparently faked results on morality and cognition. Karen Ruggiero at the University of Texas was also found to have **falsified** a number of her results (proving that bad behaviour does not have a gender bias). Each of these psychologists—and there are quite a few more examples—was believed to have faked data. Subsequently, they all were disgraced and lost their jobs.

Another reason for non-replication is that, in studies with small **sample sizes**, statistically significant results may often be the result of chance. For example, if you ask five people if they believe that aliens from other planets visit Earth and regularly abduct humans, you may get three people who agree with this notion—simply by chance. Their answers may, in fact, not be at all representative of the larger population. On the other hand, if you survey 1,000 people, there is a higher probability that their belief in alien abductions reflects the actual attitudes of society. Now, consider this scenario in the context of replication: if you try to replicate the first study—the one in which you interviewed only five people—there is only a small chance that you will randomly draw five new people with exactly the same (or similar) attitudes. It is far more likely that you will be able to replicate the findings using another large sample because it is simply more likely that the findings are accurate.

Another reason for non-replication is that, while the findings in an original study may be true, they may only be true for some people in some circumstances and not necessarily universal or enduring. Imagine that a survey in the 1950s found a strong majority of respondents trust government officials. Now, imagine the same

survey administered today, with vastly different results. This example of non-replication does not invalidate the original results. Rather, it suggests that attitudes have shifted over time.

A final reason for non-replication relates to the quality of the replication rather than the quality of the original study. Non-replication might be the product of scientist-error, with the newer investigation not following the original procedures closely enough. Similarly, the attempted replication study might, itself, have too small a sample size or insufficient statistical power to find significant results.

In Defence of Replication Attempts

Failures in replication are not all bad and, in fact, some non-replication should be expected in science. Original studies are conducted when an answer to a question is uncertain. That is to say, scientists are venturing into new territory. In such cases, we should expect some answers to be uncovered that will not pan out in the long run. Furthermore, we hope that scientists take on challenging new topics that come with some amount of risk. After all, if scientists were only to publish safe results that were easy to replicate, we might have very boring studies that do not advance our knowledge very quickly. But, with such risks, some non-replication of results is to be expected.

A recent example of risk-taking can be seen in the research of social psychologist Daryl Bem. In 2011, Bem published an article claiming he had found a number of studies stating that future events could influence the past. His proposition turns the nature of time, which is assumed by virtually everyone except science fiction writers to run in one direction, on its head. Needless to say, attacks on Bem's article came fast and furious, including attacks on his statistics and methodology (Ritchie et al., 2012). There were attempts at replication and most of them failed, but not all. A year

after Bem's article came out, the prestigious journal where it was published, *Journal of Personality and Social Psychology*, published another paper in which a scientist failed to replicate Bem's findings in a number of studies very similar to the originals (Galak et al., 2012).

Some people viewed the publication of Bem's (2011) original study as a failure in the system of science. They argued that the paper should not have been published. But the editor and reviewers of the article had moved forward with publication because, although they might have thought the findings provocative and unlikely, they did not see obvious flaws in the methodology. We see the publication of Bem's paper, and the ensuing debate, as a strength of science. We are willing to consider unusual ideas if there is evidence to support them: we are open-minded. At the same time, we are critical and believe in replication. Scientists should be willing to consider unusual or risky hypotheses but, ultimately, allow good evidence to have the final say, not people's opinions.

Solutions to the Problem

Dissemination of Replication Attempts

Here are some resources known for dissemination replication attempts:

- Psychfiledrawer.org: Archives attempted replications of specific studies and whether replication was achieved.
- Center for Open Science: Psychologist Brian Nosek, a champion of replication in psychology, has created the Open Science Framework, where replications can be reported.
- Association of Psychological Science: Has registered replications of studies, with the overall results published in *Perspectives on Psychological Science*.
- Plos One (Public Library of Science): Publishes a broad range of articles, including failed replications, and occasional summaries of replication attempts in specific areas.
- Replicability-Index: Created in 2014 by Ulrich Schimmack, the so-called "R-Index" is a statistical tool for estimating the replicability of studies, journals, and even specific researchers. Schimmack describes it as a "doping test".

The fact that replications, including failed replication attempts, now have outlets where they can be communicated to other researchers is a very encouraging development and should strengthen the science considerably. One problem for many decades has been the near impossibility of publishing replication attempts, regardless of whether they were positive or negative.

The fact that replications, including failed replication attempts, now have outlets where they can be communicated to other researchers is a very encouraging development, and should

strengthen the science considerably. One problem for many decades has been the near-impossibility of publishing replication attempts, regardless of whether they've been positive or negative.

Figure 8.1 *6 Principles of Open Science – adapted from openscienceASAP*
Jessica/Feb 20: Missing license of image

The reward structure in academia has served to discourage replication. Many psychologists—especially those who work full-time at universities—are often rewarded at work—with promotions, pay raises, tenure, and prestige—through their research. Replications of one's own earlier work, or the work of others, is typically discouraged because it does not represent original thinking. Instead, academics are rewarded for high numbers of publications, and flashy studies are often given prominence in media reports of published studies.

Psychological scientists need to carefully pursue programmatic research. Findings from a single study are rarely adequate and should be followed up by additional studies using varying

methodologies. Thinking about research this way—as if it were a program rather than a single study—can help. We would recommend that laboratories conduct careful sets of interlocking studies, where important findings are followed up using various methods. It is not sufficient to find some surprising outcome, report it, and then move on. When findings are important enough to be published, they are often important enough to prompt further, more conclusive research. In this way, scientists will discover whether their findings are replicable and how broadly generalizable they are. If the findings do not always replicate, but do sometimes, we will learn the conditions in which the pattern does or does not hold up. This is an important part of science—to discover how generalizable the findings are.

When researchers criticize others for being unable to replicate the original findings—saying that the conditions in the follow-up study were changed—is important to pay attention to, as well. Not all criticism is knee-jerk defensiveness or resentment. The replication crisis has stirred heated emotions among research psychologists and the public, but it is time for us to calm down and return to a more scientific attitude and system of programmatic research.

Open Science Practices

It is important to shed light on questionable research practices to ensure that current and future researchers (such as yourself) understand the damage they wreak to the integrity and reputation of our discipline (see, for example, the "Replicability-Index," a statistical "doping test" developed by Ulrich Schimmack in 2014 for estimating the replicability of studies, journals, and even specific researchers). However, in addition to highlighting *what not to do*, this so-called "crisis" has also highlighted the importance of enhancing scientific rigour by:

1. Designing and conducting studies that have sufficient statistical power in order to increase the reliability of findings.
2. Publishing both null and significant findings (thereby counteracting the publication bias and reducing the file drawer problem).
3. Describing one's research designs in sufficient detail to enable other researchers to replicate the study using an identical or at least very similar procedure.
4. Conducting high-quality replications and publishing these results. (Brandt et al., 2014)

One particularly promising response to the replicability crisis has been the emergence of **open science practices** that increase the transparency and openness of the scientific enterprise. For example, *Psychological Science* (the flagship journal of the Association for Psychological Science) and other journals now issue digital badges to researchers who pre-registered their hypotheses and data analysis plans, openly shared their research materials with other researchers (e.g., to enable attempts at replication), or made their raw data available to other researchers (see Figure 10.2).

Figure 8.2 *Digital Badges from the Center for Open Science (Center for Open Science) Jessica/Feb 20: Missing license of image*

These initiatives, which have been spearheaded by the Center for Open Science, have led to the development of "Transparency and

Openness Promotion guidelines" (see Table 10.2) that have since been formally adopted by more than 500 journals and 50 organizations, a list that grows each week (Nosek et al., 2015). When you add to this the requirements recently imposed by federal funding agencies in Canada (the Tri-Council) and the United States (National Science Foundation) concerning the publication of publicly-funded research in open access journals, it certainly appears that the future of science and psychology will be one that embraces greater "openness."

Table 10.2 Transparency and Openness Promotion (TOP) Guidelines, 2015

Criteria	Level 0	Level 1	Level 2	Level 3
Citation standards	Journal encourages citation of data, code, and materials, or says nothing	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used consistent with journal's author guidelines.	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.
Data transparency	Journal encourages data sharing, or says nothing	Article states whether data are available, and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Analytic methods (Code) transparency	Journal encourages code sharing, or says nothing	Article states whether code is available, and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Research materials transparency	Journal encourages materials sharing, or says nothing	Article states whether materials are available, and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.

Design and analysis transparency	Journal encourages design and analysis transparency, or says nothing	Journal articulates design transparency standards	Journal requires adherence to design transparency standards for review and publication	Journal requires and enforces adherence to design transparency standards for review and publication
Preregistration of studies	Journal says nothing	Journal encourages preregistration of studies and provides link in article to preregistration if it exists	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing	Journal encourages pre-analysis plans and provides link in article to registered analysis plan if it exists	Journal encourages pre-analysis plans and provides link in article and certification of meeting registered analysis plan badge requirements	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies, or says nothing	Journal encourages submission of replication studies	Journal encourages submission of replication studies and conducts results blind review	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.

Note. Table reproduced with permission from Center for Open Science (Nosek et al., 2015)

Textbooks and Journals

Some psychologists blame the trend toward non-replication on specific journal policies, such as the policy of *Psychological Science* to publish short single studies. When single studies are published, we do not know whether even the authors themselves can replicate their findings. The journal *Psychological Science* has come under perhaps the harshest criticism. Others blame the rash of nonreplicable studies on a tendency of some fields for surprising and counterintuitive findings that grab the public interest. The irony here is that such counterintuitive findings are, in fact, less likely to be true precisely because they are so strange—so they should perhaps warrant *more* scrutiny and further analysis.

The criticism of journals extends to textbooks, as well. In our opinion, psychology textbooks should stress true science, based on findings that have been demonstrated to be replicable. There are a number of inaccuracies that persist across common psychology textbooks, including small mistakes in common coverage of the most famous studies, such as the Stanford Prison Experiment (Griggs & Whitehead, 2014) and the Milgram studies (Griggs & Whitehead, 2015). To some extent, the inclusion of non-replicated studies in textbooks is the product of market forces. Textbook publishers are under pressure to release new editions of their books, often far more frequently than advances in psychological science truly justify. As a result, there is pressure to include "sexier" topics such as controversial studies.

Ultimately, people also need to learn to be intelligent consumers of science. Instead of getting overly excited by findings from a single study, it is wise to wait for replications. When a corpus of studies is built on a phenomenon, we can begin to trust the findings. Journalists must also be educated about this and learn not to readily broadcast and promote findings from single flashy studies. If the results of a study seem too good to be true, maybe they are.

Everyone needs to take a more skeptical view of scientific findings, until they have been replicated.

Media Attributions

Jessica/Feb 20: Media attributions might need to get fixed

- **Figure 8.1** openscienceASAP. Underlying Image: CC by-SA 3.0; Greg Emmerich)
- **Figure 8.2** Center for open Science

Summary

Key Takeaways

- The replication of findings is an important hallmark of science.
- Replication projects in psychology have had mixed results, with many studies not being replicated. There are multiple reasons why study results might not be replicable.

Acknowledgements

This chapter is adapted from the following sources:

The chapter "From the "Replicability Crisis" to Open Science Practices" in *Research Methods in Psychology – 2nd Canadian Edition* by Chiang et al. (2015), via BCcampus, is adapted under a CC BY-NC-SA 4.0 license.

The chapter "2.3 Analyzing Findings" in *Psychology 2e* by Spielman et al. (2020), via OpenStax, is used under a CC BY license.

References

Aarts, A. A., Anderson, C. J., Anderson, J., van Assen, M. A. L. M., Attridge, P. R., Attwood, A. S., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowen, M., Bartmess, E., Beer, J., Bell, R., Bentley, D., von den Bergh, D., Beyan, L., den Bezemer, B., Borsboom, D.,... Zuni, K. (2015, September 21). *Reproducibility Project: Psychology*. https://doi.org/10.17605/OSF.IO/EZCUJ

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, 193(5), 31–35. https://www.jstor.org/stable/24943779

Aschwanden, C. (2015, August 19). Science isn't broken: It's just a hell of a lot harder than we give it credit for. *Fivethirtyeight*. http://fivethirtyeight.com/features/science-isnt-broken/

Aschwanden, C. (2018, December 6). Psychology's replication crisis has made the field better. *FiveThirtyEight*. https://fivethirtyeight.com/features/psychologys-replication-crisis-has-made-the-field-better/

BBC. (2020, January 3). *Nobel Prize-winning scientist Frances Arnold retracts paper*. https://www.bbc.com/news/world-us-canada-50989423

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407–425. https://doi.org/10.1037/a0021524

Biswas-Diener, R., & Diener, E. (2001). Making the best of a bad situation: Satisfaction in the slums of Calcutta. *Social Indicators Research*, 55, 329–352. https://doi.org/10.1023/A:1010905029386

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224. https://doi.org/10.1016/j.jesp.2013.10.005

Chiang, I-C., A., Jhangiani, R. S., & Price, P. C. (2015). *Research*

methods in psychology – 2nd Canadian edition. BCcampus. https://opentextbc.ca/researchmethods/

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology*, 74(4), 865–77. https://doi.org/10.1037/0022-3514.74.4.865

Frank, M. (2015, August 31). *The slower, harder ways to increase reproducibility*. *Babies Learning Language*. http://babieslearninglanguage.blogspot.ie/2015/08/the-slower-harder-ways-to-increase.html

Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103(6), 933–948. https://doi.org/10.1037/a0029709

Griggs, R. A., & Whitehead, G. I. III. (2014). Coverage of the Stanford prison experiment in introductory social psychology textbooks. *Teaching of Psychology*, 41(4), 318–324. https://doi.org/10.1177/0098628314549703

Griggs, R. A., & Whitehead, G. I. III. (2015). Coverage of Milgram's obedience experiments in social psychology textbooks: Where have all the criticisms gone? *Teaching of Psychology*, 42(4), 315–322. https://doi.org/10.1177/0098628315603065

Head M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of *p*-hacking in science. *PLoS Biology*, 13(3), Article e1002106. https://doi.org/10.1371/journal.pbio.1002106

Kahneman, D. (2012). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni,

T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Offit, P. A. (2008). *Autism's False Prophets: Bad Science, Risky Medicine, and the Search for a Cure*. Columbia University Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments explained. *Perspectives on Psychological Science*, 7(6), 531–536. https://doi.org/10.1177/1745691612463401

Patel, R., Ashcroft, J., Patel, A., Ashrafian, H., Woods, A. J., Singh, H., Darzi, A., & Leff, D. R. (2019). The impact of transcranial direct current stimulation on upper-limb motor performance in healthy adults: A systematic review and meta-analysis. *Frontiers Neuroscience*, 13, Article 1213. https://doi.org/10.3389/fnins.2019.01213

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'retroactive facilitation of recall' effect. *PLoS ONE*, 7(3), Article e33423. https://doi.org/10.1371/journal.pone.0033423

Scherer, L. (2015, September). *Guest post by Laura Scherer. Sometimes I'm Wrong*. http://sometimesimwrong.typepad.com/wrong/2015/09/guest-post-by-laura-scherer.html

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., Kavvadia, F., & Moore, C. (2013). Priming intelligent behavior: an elusive phenomenon. *PloS one*, 8(4), Article e56515. https://doi.org/10.1371/journal.pone.0056515

Spielman, R. M., Jenkins, W. J., & Lovett, M. D. (2020). *Psychology 2e*. OpenStax. https://openstax.org/books/psychology-2e/pages/1-introduction

Spinoza, Baruch, *Theological-Political Treatise*, ed. Jonathan Israel, trans. Michael Silverthone and Jonathan Israel (London: Cambridge University Press, 2007).

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth

promotes interpersonal warmth. *Science*, 322(5901), 606–607. https://doi.org/10.1126/science.1162548

XI. CULT CLASSICS

Learning Objectives

- Describe some of the active and passive ways that conformity occurs in our everyday lives.
- Compare and contrast informational social influence and normative social influence.
- Summarize the variables that create majority and minority social influence.
- Outline the situational variables that influence the extent to which we conform.
- Describe and interpret the results of Stanley Milgram's research on obedience to authority.
- Compare the different types of power proposed by John French and Bertram Raven and explain how they produce conformity.

Introduction

"The people of England think themselves free but they choose what is customary in preference to their inclination until it does not occur to them to have any inclination except for what is customary."

— John Stuart Mill, *On Liberty*

Think about all the things you believe in. How many of those beliefs are actually your own and how many are due to the influence

of others? Even things you think of as scientific fact (such as the Earth orbiting the sun or the universe being made of atoms) are not verified facts that you discovered; they are beliefs you have because other people have told you about them. We have some confidence in their veracity because they have been through the scientific process of peer review and replication (or multiple independent observations). But what about other ideas that come to you through social conformity and group conformity? This chapter will look into some of those influences and the factors that can help us fight against such influences.

The Many Varieties of Conformity

The typical outcome of social influence is that our beliefs and behaviours become more similar to those of others around us. At times, this change occurs in a spontaneous and automatic sense, without any obvious intent of one person to change the other. Perhaps you learned to like jazz or rap music because your roommate was playing a lot of it. You did not really want to like the music, and your roommate did not force it on you—your preferences changed in a passive way.

Robert Cialdini and his colleagues (1990) found that college students were more likely to throw litter on the ground when they had just seen another person throw some paper on the ground and were less likely to litter when they had just seen another person pick up and throw paper into a trash can. The researchers interpreted this as a kind of spontaneous conformity—a tendency to follow the behaviour of others, often entirely outside of our awareness. Even our emotional states become more similar to those we spend more time with (Anderson et al., 2003).

Imitation as Subtle Conformity

Perhaps you have noticed a type of very subtle conformity—the tendency to imitate other people around you—in your own behaviour. Have you ever found yourself talking, smiling, or frowning in the same way that a friend does? Tanya Chartrand and John Bargh (1999) investigated

whether the tendency to imitate others would occur even for strangers, and even in very short periods of time.

In their first experiment, students worked on a task with another student, who was actually an experimental confederate. The two worked together to discuss photographs taken from current magazines. While they were working together, the confederate engaged in some unusual behaviours to see if the research participant would mimic them. Specifically, the confederate either rubbed his or her face or shook his or her foot. It turned out that the students did mimic the behaviour of the confederate by either rubbing their own faces or shaking their own feet. When the experimenters asked the participants if they had noticed anything unusual about the behaviour of the other person during the experiment, none of them indicated awareness of any face rubbing or foot shaking.

It is said that imitation is a form of flattery, and, therefore, we might expect that we would like people who imitate us. Indeed, in a second experiment, Chartrand and Bargh (1999) found exactly this. Rather than creating the behaviour to be mimicked, in this study, the confederate imitated the behaviours of the participant. While the participant and the confederate discussed the magazine photos, the confederate mirrored the posture, movements, and mannerisms displayed by the participant.

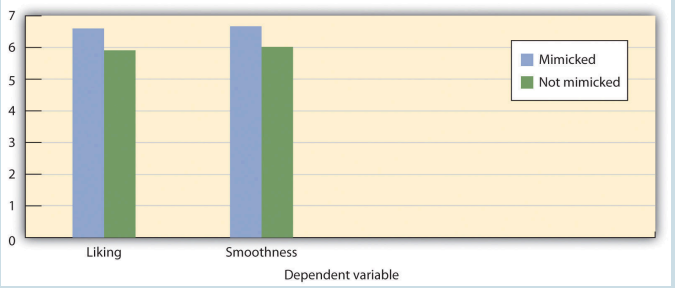


Figure 9.1 *Mimicking and liking (Jhangiani & Tarry, 2022)*

As you can see in Figure 9.1, the participants who had been mimicked liked the other person more and indicated that they thought the interaction had gone more smoothly, in comparison with the participants who had not been imitated.

Informational Social Influence: Conforming to Be Accurate

Although mimicry represents the more subtle side, social influence also occurs in a more active and thoughtful sense, such as when we actively look to our friends' opinions to determine appropriate behaviour, when a car salesperson attempts to make a sale, or even when a powerful dictator uses physical aggression to force the people in his country to engage in the behaviours that he desires.

In these cases, the influence is obvious. We know we are being influenced, and we may attempt—sometimes successfully, and sometimes less so—to counteract the pressure.

Influence also sometimes occurs because we believe that other people have valid knowledge about an opinion or issue, and we use that information to help us make good decisions. For example, if you take a flight and land at an unfamiliar airport, you may follow the flow of other passengers who disembarked before you. In this case, your assumption might be that they know where they are going and that following them will likely lead you to the baggage carousel.

Informational social influence is the change in opinions or behaviour that occurs when we conform to people who we believe have accurate information. We base our beliefs on those presented to us by reporters, scientists, doctors, and lawyers because we believe they have more expertise in certain fields than we have. But we also use our friends and colleagues for information; when we choose a jacket on the basis of our friends' advice about what looks good on us, we are using informational conformity—we believe that our friends have good judgment about the things that matter to us.

Informational social influence is often the end result of social comparison—the process of comparing our opinions with those of others to gain an accurate appraisal of the validity of an opinion or behaviour (Festinger et al., 1950; Hardin & Higgins, 1996; Turner, 1991). Informational social influence leads to real, long-lasting changes in beliefs. The result of conformity due to informational social influence is normally private acceptance: real change in opinions on the part of the individual. We believe that choosing the jacket was the right thing to do and that the crowd will lead us to the baggage carousel.

Normative Social Influence: Conforming to Be Liked and to Avoid Rejection

In other cases, we conform not because we want to have valid knowledge but rather to meet the goal of belonging to and being accepted by a group that we care about (Deutsch & Gerard, 1955). When we start smoking cigarettes or buy shoes that we cannot really afford in order to impress others, we do these things not so much because we think they are the right things to do but rather because we want to be liked.

We fall prey to normative social influence when we express opinions or behave in ways that help us be accepted or that keep us from being isolated or rejected by others. When we engage in conformity due to normative social influence, we conform to social norms—socially accepted beliefs about what we do or should do in particular social contexts (Cialdini, 1993; Sherif, 1936; Sumner, 1906).

In contrast to informational social influence—in which the attitudes or opinions of the individual change to match that of the influencers—the outcome of normative social influence often represents public compliance rather than private acceptance. Public compliance is a superficial change in behaviour (including the public expression of opinions) that is not accompanied by an actual change in one's private opinion. Conformity may appear in our public behaviour even though we may believe something completely different in private. We may obey the speed limit or wear a uniform to our job (behaviour) to conform to social norms and requirements, even though we may not necessarily believe that it is appropriate to do so (opinion). We may use drugs with our friends without really wanting to, and without believing it is really right, because our friends are all using drugs. However, behaviours that are originally performed out of a desire to be accepted (normative social influence) may frequently produce changes in beliefs to match them, and the result becomes private acceptance. Perhaps you

know someone who started smoking to please his friends but soon convinced himself that it was an acceptable thing to do.

Although in some cases conformity may be purely informational or purely normative, in most cases, the goals of being accurate and being accepted go hand-in-hand, and therefore, informational and normative social influence often occur at the same time. When soldiers obey their commanding officers, they probably do it both because others are doing it (normative conformity) and because they think it is the right thing to do (informational conformity). And when you start working at a new job, you may copy the behaviour of your new colleagues because you want them to like you as well as because you assume they know how things should be done. It has been argued that the distinction between informational and normative conformity is more apparent than real and that it may not be possible to fully differentiate them (Turner, 1991).

Majority Influence: Conforming to the Group

Although conformity occurs whenever group members change their opinions or behaviours as a result of their perceptions of others, we can divide such influence into two types. Majority influence occurs when the beliefs held by the larger number of individuals in the current social group prevail. In contrast, minority influence occurs when the beliefs held by the smaller number of individuals in the current social group prevail. Not surprisingly, majority influence is more common, and we will consider it first.

In a series of important studies on conformity, Muzafer Sherif (1936) used a perceptual phenomenon known as the autokinetic effect to study the outcomes of conformity on the development of group norms. The autokinetic effect is caused by the rapid, small movements of our eyes that occur as we view objects and that allow us to focus on stimuli in our environment. However, when individuals are placed in a dark room that contains only a single

small, stationary pinpoint of light, these eye movements produce an unusual effect for the perceiver—they make the point of light appear to move.

Sherif (1936) took advantage of this natural effect to study how group norms develop in ambiguous situations. In his studies, college students were placed in a dark room with the point of light and were asked to indicate, each time the light was turned on, how much it appeared to move. Some participants first made their judgments alone. Sherif found that although each participant who was tested alone made estimates that were within a relatively narrow range (as if they had their own "individual" norm), there were wide variations in the size of these judgments among the different participants he studied.

Sherif (1936) also found that when individuals who initially had made very different estimates were placed in groups along with one or two other individuals, and in which all the group members gave their responses on each trial aloud (each time in a different random order), the initial differences in judgments among the participants began to disappear, such that the group members eventually made very similar judgments. You can see that this pattern of change, which is shown in Figure 9.2, "Outcomes of Sherif's Study," illustrates the fundamental principle of social influence—over time, people share their beliefs more and more with each other. Thus, Sherif's study is a powerful example of the development of group norms.

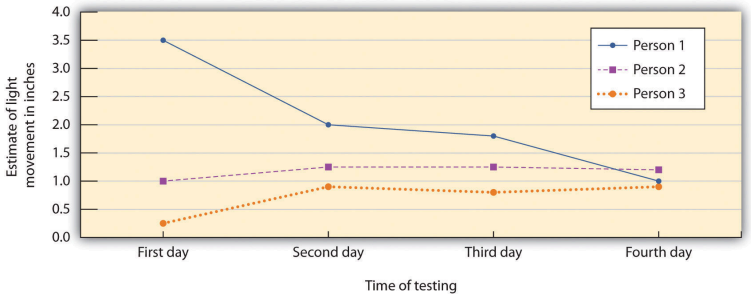


Figure 9.2 *Outcomes of Sherif's Study (Jhangiani & Tarry, 2022)*

The participants in the studies by Muzafer Sherif (1936) initially had different beliefs about the degree to which a point of light appeared to be moving. (You can see these differences as expressed on Day 1.) However, as they shared their beliefs with other group members over several days, a common group norm developed. Shown here are the estimates made by a group of three participants who met together on four different days.

Furthermore, the new group norms continued to influence judgments when the individuals were again tested alone, indicating that Sherif had created private acceptance (Sherif, 1936). The participants did not revert back to their initial opinions, even though they were quite free to do so; rather, they stayed with the new group norms. And these conformity effects appear to have occurred entirely out of the awareness of most participants. Sherif reported that the majority of the participants indicated after the experiment was over that their judgments had not been influenced by the judgments made by the other group members.

Sherif (1936) also found that the norms developed in groups could continue over time. When the original research participants were moved into groups with new people, their opinions subsequently influenced the judgments of the new group members (Jacobs & Campbell, 1961). The norms persisted through several "generations" (MacNeil & Sherif, 1976) and could influence individual judgments up to a year after the individual was last tested (Rohrer et al., 1954).

When Solomon Asch (1952, 1955) heard about Sherif's studies, he responded in perhaps the same way that you might have: "Well of course people conformed in this situation, because after all the right answer was very unclear," you might have thought. Since the study participants did not know the right answer (or indeed the "right" answer was no movement at all), it is perhaps not that surprising that people conformed to the beliefs of others.

Asch (1952, 1955) conducted studies in which, in complete contrast to the autokinetic effect experiments of Sherif, the correct answers to the judgments were entirely obvious. In these studies, the research participants were male college students who were

told that they were to be participating in a test of visual abilities. The men were seated in a small semicircle in front of a board that displayed the visual stimuli that they were going to judge. The men were told that there would be 18 trials during the experiment, and on each trial, they would see two cards. The standard card had a single line that was to be judged. And the test card had three lines that varied in length between about 2 and 10 inches.

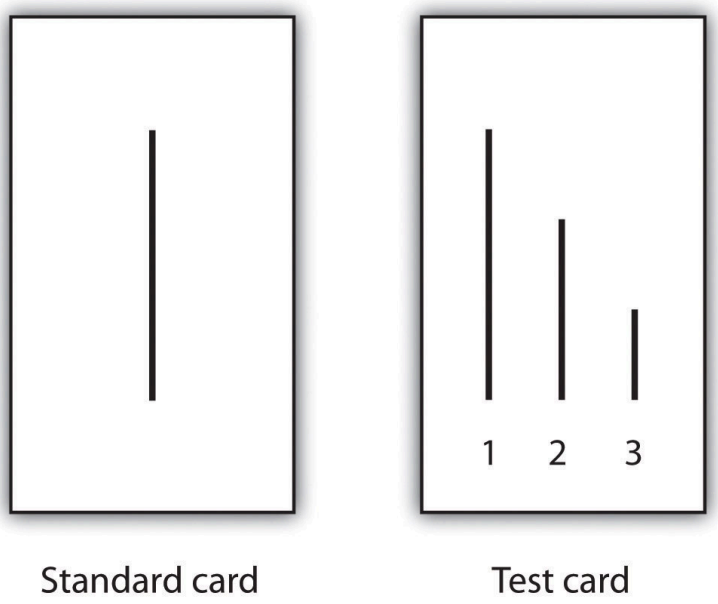


Figure 9.3 *Standard Card and Test Card (Jhangiani & Tarry, 2022)*

The men's task was simply to indicate which line on the test card was the same length as the line on the standard card. As you can see from the Asch (1952, 1955) card sample above, there is no question that correct answer is Line 1. In fact, Asch found that people made virtually no errors on the task when they made their judgments alone.

On each trial, each person answered out loud, beginning with one

end of the semicircle and moving to the other end (Asch, 1952, 1955). Although the participant did not know it, the other group members were not true participants but experimental confederates who gave predetermined answers on each trial. Because the participant was seated next to last in the row, he always made his judgment after most of the other group members made theirs. Although on the first two trials the confederates each gave the correct answer, on the third trial, and on 11 of the subsequent trials, they all had been instructed to give the same incorrect answer. For instance, even though the correct answer was Line 1, they would all say it was Line 2. Thus, when it became the participant's turn to answer, he could either give the clearly correct answer or conform to the incorrect responses of the confederates.

Asch (Asch, 1952, 1955) found that about 76% of the 123 men who were tested gave at least one incorrect response when it was their turn, and 37% of the responses, overall, were conforming. This is indeed evidence for the power of normative social influence because the research participants were giving clearly incorrect answers out loud. However, conformity was not absolute—in addition to the 24% of the men who never conformed, only 5% of the men conformed on all 12 of the critical trials.

Minority Influence: Resisting Group Pressure

The research that we have discussed to this point involves conformity in which the opinions and behaviours of individuals become more similar to the opinions and behaviours of the majority of the people in the group—this is majority influence. But we do not always blindly conform to the beliefs of the majority. Although more unusual, there are nevertheless cases in which a smaller number of individuals are able to influence the opinions or behaviours of the group—this is minority influence.

It is a good thing that minorities can be influential; otherwise,

the world would be pretty boring. When we look back on history, we find that it is the unusual, divergent, innovative minority groups or individuals, who—although frequently ridiculed at the time for their unusual ideas—end up being respected for producing positive changes. The work of scientists, religious leaders, philosophers, writers, musicians, and artists who go against group norms by expressing new and unusual ideas are frequently not liked at first. Galileo and Copernicus were scientists who did not conform to the opinions and behaviours of those around them. In the end, their innovative ideas changed the thinking of the masses. These novel thinkers may be punished—in some cases, even killed—for their beliefs. In the end, however, if the ideas are interesting and important, the majority may conform to these new ideas, producing social change. In short, although conformity to majority opinions is essential to provide a smoothly working society, if individuals only conformed to others there would be few new ideas and little social change.

The French social psychologist Serge Moscovici was particularly interested in the situations under which minority influence might occur. In fact, he argued that all members of all groups are able, at least in some degree, to influence others, regardless of whether they are in the majority or the minority. To test whether minority group members could indeed produce influence, Moscovici and his colleagues (1969) created the opposite of Asch's line perception study, such that there was now a minority of confederates in the group (two) and a majority of experimental participants (four).

All six individuals viewed a series of slides depicting colours, supposedly as a study of colour perception, and as in Asch's research, each voiced out loud an opinion about the color of the slide (Moscovici et al., 1969). Although the colour of the slides varied in brightness, they were all clearly. Moreover, demonstrating that the slides were unambiguous—just as the line judgments of Asch had been—participants who were asked to make their judgments alone called the slides a different colour than blue less than 1% of the time. (When it happened, they called the slides green.)

In the experiment, the two confederates had been instructed to give one of two patterns of answers that were different from the normal responses (Moscovici et al., 1969). In the consistent-minority condition, the two confederates gave the unusual response (green) on every trial. In the inconsistent-minority condition, the confederates called the slides green on two-thirds of their responses and called them blue on the other third.

The minority of two was able to change the beliefs of the majority of four, but only when they were unanimous in their judgments (Moscovici et al., 1969). As shown in Figure 9.4, "The Power of Consistent Minorities," Moscovici and his colleagues found that the presence of a minority who gave consistently unusual responses influenced the judgments made by the experimental participants. When the minority was consistent, 32% of the majority group participants said green at least once and 18% of the responses of the majority group were green. However, the inconsistent minority had virtually no influence on the judgments of the majority.

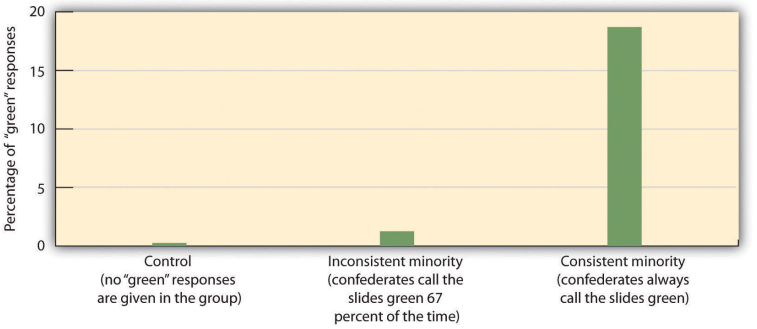


Figure 9.4 *The Power of Consistent Minorities (Jhangiani & Tarry, 2022)*

In the studies of minority influence by Serge Moscovici, only a consistent minority (in which each individual gave the same incorrect response) was able to produce conformity in the majority participants (Moscovici, Lage, and Naffrechoux, 1969). On the basis of this research, Moscovici argued that minorities could have

influence over majorities, provided they gave consistent, unanimous responses. Subsequent research has found that minorities are most effective when they express consistent opinions over time and with each other, when they show that they are invested in their position by making significant personal and material sacrifices, and when they seem to be acting out of principle rather than from ulterior motives (Hogg, 2010). Although they may want to adopt a relatively open-minded and reasonable negotiating style on issues that are less critical to the attitudes they are trying to change, successful minorities must be absolutely consistent with their core arguments (Mugny & Papastamou, 1981).

When minorities are successful at producing influence, they are able to produce strong and lasting attitude change—true private acceptance—rather than simply public compliance. People conform to minorities because they think that they are right, and not because they think it is socially acceptable. Minorities have another, potentially even more important, outcome on the opinions of majority group members—the presence of minority groups can lead majorities to engage in fuller, as well as more divergent, innovative and creative thinking about the topics being discussed (Martin & Hewstone, 2003; Martin et al., 2007).

Nemeth and Kwan (1987) had participants work in groups of four on a creativity task in which they were presented with letter strings such as "tdogto" and asked to indicate which word came to their mind first as they looked at the letters. The judgments were made privately, which allowed the experimenters to provide false feedback about the responses of the other group members. All participants indicated the most obvious word (in this case, "dog") as their response on each of the initial trials. However, the participants were told (according to experimental condition) either that three of the other group members had also reported seeing "dog" and one had reported seeing "god" or that three out of the four had reported seeing "god" whereas only one had reported "dog." Participants then completed other similar word strings on their own, and their responses were studied.

Results showed that when the participants thought that the unusual response (for instance, "god" rather than "dog") was given by a minority of one individual in the group rather than by a majority of three individuals, they subsequently answered more of the new word strings using novel solutions, such as finding words made backwards or using a random order of the letters (Nemeth & Kwan, 1987). On the other hand, the individuals who thought that the majority of the group had given the novel response did not develop more creative ideas. Evidently, when the participants thought that the novel response came from a group minority (one person), they thought about the responses more carefully, in comparison with the same behaviours performed by majority group members; this led them to adopt new and creative ways to think about the problems. This result, along with other research showing similar findings, suggests that messages that come from minority groups lead us to think more fully about the decision, which can produce innovative, creative thinking in majority group members (Crano & Chen, 1998).

The Fool Who Would Challenge the King

9.5 *Ceramic Sacred Clown*

Then they for sudden joy did weep

And I for sorrow sung,

That such a king should play bo-peep

And go the fools among.

Prithee, nuncle, keep a schoolmaster that can teach thy fool to lie. I would fain learn to lie.

– *King Lear*, Shakespeare

The above statement from *King Lear* is an example of a court jester or fool speaking truth to power. Such a character is in the minority but their presence allows the others to have some insight which might have escaped them due to conformity or groupthink.

Some cultures have similar figures as a stopgap against group conformity. Cultures in India have the *Vikatakavi* or humorous poet who enlightens a king through their clownish songs and behaviour. Some historical poets have become legends in modern India (Birbal who served the Mughal emperor Akbar or *Tenaliramakrishna* who served the emperor *Krishnadevraya*). Such figures are also found in the religious realm as *Avadhuta* or messenger who shakes (preconceptions).

The Sioux (Lakota and Dakota people) of the Great Plains of North America have the *Heyoka* who serve as sacred clowns. Only people who have had dreams of the thunder beings (or *Wakingyang*) can become *Heyoka*. The role of this sacred jester is to go against the majority. He will always do the opposite of what the tribe is doing as a way to remind people of other alternatives and possibilities. Having such a figure who can challenge the chiefs and elders with impunity was a useful mechanism that allowed a tribe to have immunity from groupthink. Similar figures are found in what is now the Southwestern United States in the form of Pueblo Clowns. Their role is the same: to critique and make fun of defective aspects of their own culture.

In summary, we can conclude that minority influence, although not as likely as majority influence, does sometimes occur. The few are able to influence the many when they are consistent and confident in their judgments but are less able to have influence when they are inconsistent or act in a less confident manner. Furthermore, although minority influence is difficult to achieve, if it does occur it is powerful. When majorities are influenced by minorities, they really change their beliefs—the outcome is deeper thinking about the message, private acceptance of the message, and in some cases even more creative thinking.

Media Attributions

- **Figure 9.1** Jhangiani, R., & Tarry, H. (2022). Principles of social psychology (1st international H5P edition). BCcampus. https://opentextbc.ca/socialpsychology/
- **Figure 9.2** Jhangiani, R., & Tarry, H. (2022). Principles of social psychology (1st international H5P edition). BCcampus. https://opentextbc.ca/socialpsychology/
- **Figure 9.3** Jhangiani, R., & Tarry, H. (2022). Principles of social psychology (1st international H5P edition). BCcampus. https://opentextbc.ca/socialpsychology/
- **Figure 9.4** Jhangiani, R., & Tarry, H. (2022). Principles of social psychology (1st international H5P edition). BCcampus. https://opentextbc.ca/socialpsychology/
- **Figure 9.5** Fisher, M. (2011) Kathleen Wall (Jemez) ceramic Koshare sculpture https://www.flickr.com/photos/bfs_man/6598035473/

Obedience and Power

One of the fundamental aspects of social interaction is that some individuals have more influence than others. Social power can be defined as the ability of a person to create conformity even when the people being influenced may attempt to resist those changes (Fiske, 1993; Keltner et al., 2003). Bosses have power over their workers, parents have power over their children, and, more generally, we can say that those in authority have power over their subordinates. In short, power refers to the process of social influence itself—those who have power are those who are most able to influence others.

Milgram's Studies on Obedience to Authority

The powerful ability of those in authority to control others was demonstrated in a remarkable set of studies performed by Stanley Milgram (1963). Milgram was interested in understanding the factors that lead people to obey the orders given by people in authority. He designed a study in which he could observe the extent to which a person who presented himself as an authority would be able to produce obedience, even to the extent of leading people to cause harm to others.

Like his professor Solomon Asch, Milgram's interest in social influence stemmed in part from his desire to understand how the presence of a powerful person—particularly the German dictator Adolf Hitler who ordered the killing of millions of people during World War II—could produce obedience. Under Hitler's direction, the German SS troops oversaw the execution of 6 million Jews as well as other "undesirables," including political and religious dissidents, homosexuals, mentally and physically disabled people, and prisoners of war.

Milgram (1963) used newspaper ads to recruit men (and in one study, women) from a wide variety of backgrounds to participate in his research. When the research participant arrived at the lab, he or she was introduced to a man who the participant believed was another research participant but who was actually an experimental confederate. The experimenter explained that the goal of the research was to study the effects of punishment on learning. After the participant and the confederate both consented to participate in the study, the researcher explained that one of them would be randomly assigned to be the teacher and the other the learner. They were each given a slip of paper and asked to open it and indicate what it said. In fact, both papers read teacher, which allowed the confederate to pretend that he had been assigned to be the learner and, thus, assure that the actual participant was always the teacher.

While the research participant (now the teacher) looked on, the learner was taken into the adjoining shock room and strapped to an electrode that was to deliver the punishment (Milgram, 1963). The experimenter explained that the teacher's job would be to sit in the control room and read a list of word pairs to the learner. After the teacher read the list once, it would be the learner's job to remember which words went together. For instance, if the word pair was "blue-sofa," the teacher would say the word "blue" on the testing trials and the learner would have to indicate which of four possible words (house, sofa, cat, or carpet) was the correct answer by pressing one of four buttons in front of him. After the experimenter gave the "teacher" a sample shock (which was said to be at 45 volts) to demonstrate that the shocks really were painful, the experiment began. The research participant first read the list of words to the learner and then began testing him on his learning.

The shock panel, as shown in Figure 9.5, "The Shock Apparatus Used in Milgram's Obedience Study," was presented in front of the teacher, and the learner was not visible in the shock room. The experimenter sat behind the teacher and explained to him that each time the learner made a mistake, the teacher was to press one of the shock switches to administer the shock. They were to begin with

the smallest possible shock (15 volts), but with each mistake, the shock was increased by one level (an additional 15 volts).

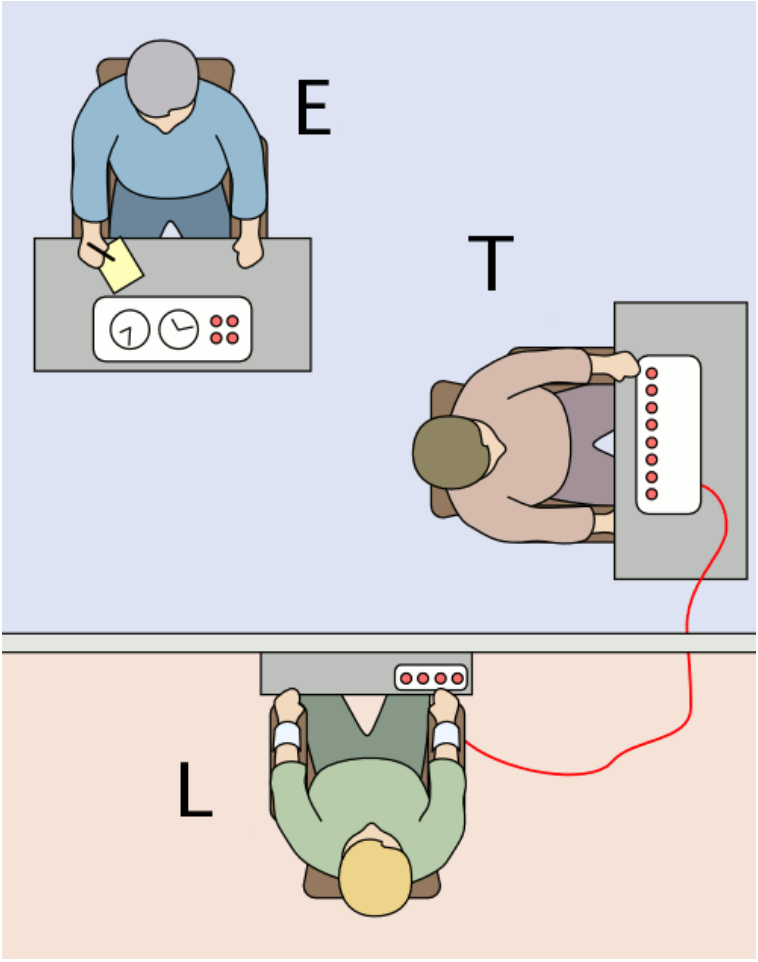


Figure 9.6 *The Shock Apparatus Used in Milgram's Obedience Study (Jhangiani & Tarry, 2022)* *Jessica/Feb 20: Image missing its license*

Once the learner (who was, of course, actually an experimental confederate) was alone in the shock room, he unstrapped himself

from the shock machine and brought out a tape recorder that he used to play a prerecorded series of responses that the teacher could hear through the wall of the room (Milgram, 1963). As you can see in Table 9.1, "The Confederate's Schedule of Protest in the Milgram Experiments," the teacher heard the learner say "ugh!" after the first few shocks. After the next few mistakes, when the shock level reached 150 volts, the learner was heard exclaiming "Get me out of here, please. My heart's starting to bother me. I refuse to go on. Let me out!" As the shock reached about 270 volts, the learner's protests became more vehement, and after 300 volts, the learner proclaimed that he was not going to answer any more questions. From 330 volts and up the learner was silent. The experimenter responded to participants' questions at this point, if they asked any, with a scripted response indicating that they should continue reading the questions and applying increasing shock when the learner did not respond.

Table 9.1 The Confederate's Schedule of Protest in the Milgram Experiments

Shock Level	Recorded Response
75 volts	Ugh!
90 volts	Ugh!
105 volts	Ugh! (louder)
120 volts	Ugh! Hey, this really hurts.
135 volts	Ugh!!
150 volts	Ugh!! Experimenter! That's all. Get me out of here. I told you I had heart trouble. My heart's starting to bother me now. Get me out of here, please. My heart's starting to bother me. I refuse to go on. Let me out!
165 volts	Ugh! Let me out! (shouting)
180 volts	Ugh! I can't stand the pain. Let me out of here! (shouting)
195 volts	Ugh! Let me out of here! Let me out of here! My heart's bothering me. Let me out of here! You have no right to keep me here! Let me out! Let me out of here! Let me out! Let me out of here! My heart's bothering me. Let me out! Let me out!
210 volts	Ugh!! Experimenter! Get me out of here. I've had enough. I won't be in the experiment any more.
225 volts	Ugh!
240 volts	Ugh!
255 volts	Ugh! Get me out of here.
270 volts	(agonized scream) Let me out of here. Let me out of here. Let me out of here. Let me out. Do you hear? Let me out of here.
285 volts	(agonized scream)

300 volts	(agonized scream) I absolutely refuse to answer any more. Get me out of here. You can't hold me here. Get me out. Get me out of here.
315 volts	(intensely agonized scream) Let me out of here. Let me out of here. My heart's bothering me. Let me out, I tell you. (hysterically) Let me out of here. Let me out of here. You have no right to hold me here. Let me out! Let me out! Let me out! Let me out of here! Let me out! Let me out!

Before Milgram (1963) conducted his study, he described the procedure to three groups—college students, middle-class adults, and psychiatrists—asking each of them if they thought they would shock a participant who made sufficient errors at the highest end of the scale (450 volts). One hundred percent of all three groups thought they would not do so. He then asked them what percentage of "other people" would be likely to use the highest end of the shock scale, at which point the three groups demonstrated remarkable consistency by all producing (rather optimistic) estimates of around 1% to 2%.

The results of the actual experiments were themselves quite shocking. Although all of the participants gave the initial mild levels of shock, responses varied after that. Some refused to continue after about 150 volts, despite the insistence of the experimenter to continue to increase the shock level. Still others, however, continued to present the questions—and administer the shocks—under the pressure of the experimenter, who demanded that they continue. In the end, 65% of the participants continued giving the shocks to the learner all the way up to the 450 volts maximum, even though that shock was marked as "danger: severe shock," and there had been no response heard from the participant for several trials. In sum, almost two-thirds of the men who participated had, as far as they knew, shocked another person to death, all as part of a supposed experiment on learning.

Studies similar to Milgram's findings have since been conducted all over the world (Blass, 1991), with obedience rates ranging from a high of 90% in Spain and the Netherlands (Meeus & Raaijmakers, 1986) to a low of 16% among Australian women (Kilham & Mann,

1974). In case you are thinking that such high levels of obedience would not be observed in today's modern culture, there is evidence that they would be. Recently, Milgram's results were almost exactly replicated, using men and women from a wide variety of ethnic groups, in a study conducted by Jerry Burger at Santa Clara University. In this replication of the Milgram experiment, 65% of the men and 73% of the women agreed to administer increasingly painful electric shocks when they were ordered to by an authority figure (Burger, 2009). In the replication, however, the participants were not allowed to go beyond the 150-volt shock switch.

Although it might be tempting to conclude that Milgram's experiments demonstrate that people are innately evil creatures who are ready to shock others to death, Milgram did not believe that this was the case. Rather, he felt that it was the social situation—and not the people themselves—that was responsible for the behaviour. To demonstrate this, Milgram conducted research that explored a number of variations on his original procedure, each of which demonstrated that changes in the situation could dramatically influence the amount of obedience. These variations are summarized in Table 9.2.

Table 9.2 Authority and Obedience in Stanley Milgram's Studies

Experimental Replication	Description	Percent Obedience
Experiment 1	Initial study: Yale University men and women	65
Experiment 10	The study is conducted off campus, in Bridgeport CT	48
Experiment 3	The teacher is in the same room as the learner	40
Experiment 4	The participant must hold the learner's hand on the shock pad	20
Experiment 7	The experimenter communicates by phone from another room	20
Experiment 13	An "ordinary man" (presumably another research participants refuse to give shock	20
Experiment 17	Two other research participants refuse to give shock	10
Experiment 11	The teacher chooses his own preferred shock level	0
Experiment 15	One experimenter indicates that the participant should not shock	0

Table 9.2 presents the percentage of participants in Stanley Milgram's (1974) studies on obedience who were maximally obedient (that is, who gave all 450 volts of shock) in some of the variations that he conducted. In the initial study, the authority's status and power was maximized—the experimenter had been introduced as a respected scientist at a respected university. However, in replications of the study in which the experimenter's authority was decreased, obedience also declined.

In one replication, the status of the experimenter was reduced by having the experiment take place in a building located in Bridgeport, Connecticut, rather than at the labs on the Yale University campus, and the research was ostensibly sponsored by a private commercial research firm instead of by the university (Milgram, 1974). In this study, less obedience was observed (only 48% of the participants delivered the maximum shock). Full obedience was also reduced (to

20%) when the experimenter's ability to express his authority was limited by having him sit in an adjoining room and communicate to the teacher by telephone. And when the experimenter left the room and had another student (actually a confederate) give the instructions for him, obedience was also reduced to 20%.

In addition to the role of authority, Milgram's (1974) studies also confirmed the role of unanimity in producing obedience. When another research participant (again an experimental confederate) began by giving the shocks but then later refused to continue and the participant was asked to take over, only 10% were obedient. And if two experimenters were present but only one proposed shocking while the other argued for stopping the shocks, all the research participants took the more benevolent advice and did not shock. But perhaps most telling were the studies in which Milgram allowed the participants to choose their own shock levels or in which one of the experimenters suggested that they should not actually use the shock machine. In these situations, there was virtually no shocking. These conditions show that people do not like to harm others, and when given a choice, they will not. On the other hand, the social situation can create powerful, and potentially deadly, social influence.

One final note about Milgram's studies: Although Milgram explicitly focused on the situational factors that led to greater obedience, these have been found to interact with certain personality characteristics (yet another example of a person-situation interaction). Specifically, authoritarianism (a tendency to prefer things to be simple rather than complex and to hold traditional values), conscientiousness (a tendency to be responsible, orderly, and dependable), and agreeableness (a tendency to be good-natured, cooperative, and trusting) are all related to higher levels of obedience, whereas higher moral reasoning (the manner in which one makes ethical judgments) and social intelligence (an ability to develop a clear perception of the situation using situational cues) both predict resistance to the demands of the authority figure (Bègue et al., 2014; Blass, 1991).

It is worth noting that although we have discussed both conformity and obedience in this chapter, they are not the same thing. While both are forms of social influence, we most often tend to conform to our peers, whereas we obey those in positions of authority. Furthermore, the pressure to conform tends to be implicit, whereas the order to obey is typically rather explicit. And finally, whereas people do not like admitting they conformed (especially via normative social influence), they will more readily point to the authority figure as the source of their actions (especially when they have done something they are embarrassed or ashamed of).

Media Attributions

Jessica/Feb 20: Media Attributions title added. attribution might need to be checked

- **Figure 9.6** Jhangiani, R., & Tarry, H. (2022). Principles of social psychology (1st international H5P edition). BCcampus. https://opentextbc.ca/socialpsychology/

Summary

This chapter has concerned the many and varied ways that social influence pervades our everyday lives. Perhaps you were surprised about the wide variety of phenomena—ranging from the unaware imitation of others to leadership to blind obedience to authority—that involve social influence. Yet because you are thinking like a social psychologist, you will realize why social influence is such an important part of our everyday life. For example, we conform to better meet the basic goals of self-concern and other-concern. Conforming helps us do better by helping us make accurate, informed decisions. And conformity helps us be accepted by those we care about.

Because you are now more aware of these factors, you will naturally pay attention to the times when you conform to or obey others and when you influence others to conform to or obey you. You will see how important—indeed, how amazing—the effects of social influence are. You will realize that almost everything we do involves social influence, or perhaps the desire to avoid being too influenced. Furthermore, you will realize (and hopefully use this knowledge to inform your everyday decisions) that social influence is sometimes an important part of societal functioning and that, at other times, social influence creates bad—indeed, horrible—outcomes.

You can use your understanding of social influence to help understand your own behaviour. Do you think you conform too much or too little? Do you think about when you do or do not conform? Are you more of a conformist or an independent thinker—and why do you prefer to be that way? What motivates you to obey the instructions of your professor? Is it expert power, coercive power, or referent power? Perhaps you will use your understanding of the power of social influence when you judge others. When you think about the behaviour of ordinary Germans

during World War II, do you now better understand how much they were influenced by the social situation?

Your understanding of social influence may also help you develop more satisfying relations with others. Because you now understand the importance of social influence, you will also understand how to make use of these powers to influence others. If you are in a leadership position, you now have a better idea about the many influence techniques available to you and better understand their likely outcomes on others.

Key Takeaways

- Social influence creates conformity.
- Influence may occur in more passive or more active ways.
- We conform both to gain accurate knowledge (informational social influence) and to avoid being rejected by others (normative social influence).
- Both majorities and minorities may create social influence, but they do so in different ways.
- The characteristics of the social situation, including the number of people in the majority and the unanimity of the majority, have a strong influence on conformity.
- Social power can be defined as the ability of a person to create conformity, even when the people being influenced may attempt to resist those changes.
- Milgram's studies on obedience demonstrated the remarkable extent to which the social situation and

people with authority have the power to create obedience.

Acknowledgements

This chapter is adapted from the following source:

- The book *Principles of Social Psychology (1st International H5P Edition)* by Jhangiani & Tarry (2022), via BCcampus, is adapted under a CC BY-NC-SA 4.0 license.

References

Allen, V. L. (1975). Social support for nonconformity. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 1–43). New York, NY: Academic Press.

Allen, V. L., & Levine, J. M. (1968). Social support, dissent and conformity. *Sociometry*, *31*(2), 138–149. https://doi.org/10.2307/2786454

Anderson, C., Keltner, D., & John, O. P. (2003). Emotional convergence between people over time. *Journal of Personality and Social Psychology*, *84*(5), 1054–1068. https://doi.org/10.1037/0022-3514.84.5.1054

Asch, S. E. (1952). *Social psychology*. Prentice-Hall.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31–35. https://www.jstor.org/stable/24943779

Baron, R. S., Vandello, J. A., & Brunsman, B. (1996). The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology*, *71*(5), 915–927. https://doi.org/10.1037/0022-3514.71.5.915

Bègue, L., Beauvois, J-L., Courbet, D., Oberlé, D., Lepage, J., & Duke, A. A. (2014). Personality predicts obedience in a Milgram paradigm. *Journal of Personality*, *83*(3), 299–306. https://doi.org/10.1111/jopy.12104

Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, *60*(3), 398–413. https://doi.org/10.1037/0022-3514.60.3.398

Boyanowsky, E. O., & Allen, V. L. (1973). Ingroup norms and self-identity as determinants of discriminatory behavior. *Journal of Personality and Social Psychology*, *25*(3), 408–418. https://doi.org/10.1037/h0034212

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, *64*(1), 1-11.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Chartrand, T. L., & Dalton, A. N. (2009). Mimicry: Its ubiquity, importance, and functionality. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 458–483). Oxford University Press.

Cialdini, R. B. (1993). *Influence: Science and practice* (3rd ed.). HarperCollins College Publishers.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

Crano, W. D., & Chen, X. (1998). The leniency contract and persistence of majority and minority influence. *Journal of Personality and Social Psychology*, 74(6), 1437–1450. https://doi.org/10.1037/0022-3514.74.6.1437

Dalton, A. N., Chartrand, T. L., & Finkel, E. J. (2010). The schema-driven chameleon: How mimicry affects executive and self-regulatory resources. *Journal of Personality and Social Psychology*, 98(4), 605–617. https://doi.org/10.1037/a0017629

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 51(3), 629–636. https://doi.org/10.1037/h0046408

Festinger, L., Schachter, S., & Back, K. (1950). *Social pressures in informal groups: A study of human factors in housing*. Harper.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, 48(6), 621–628. https://doi.org/10.1037/0003-066X.48.6.621

Hardin, C., & Higgins, T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E.

T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 3, pp. 28–84). The Guilford Press.

Hogg, M. A. (2010). Influence and leadership. In S. F. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 1166–1207). John Wiley and Sons.

Jacobs, R. C., & Campbell, D. T. (1961). The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. *Journal of Abnormal and Social Psychology*, 62, 649–658. https://doi.org/10.1037/h0044182

Jhangiani, R., & Tarry, H. (2022). *Principles of social psychology (1st international H5P ed.)*. BCcampus. https://opentextbc.ca/socialpsychology/

Kilham, W., & Mann, L. (1974). Level of destructive obedience as a function of transmitter and executant roles in the Milgram obedience paradigm. *Journal of Personality and Social Psychology*, 29, 692–702.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, 110(2), 265–284. https://doi.org/10.1037/0033-295X.110.2.265

Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36(4), 343–356. https://doi.org/10.1037/0003-066X.36.4.343

MacNeil, M. K., & Sherif, M. (1976). Norm change over subject generations as a function of arbitrariness of prescribed norms. *Journal of Personality and Social Psychology*, 34(5), 762–773. https://doi.org/10.1037/0022-3514.34.5.762

Martin, R., & Hewstone, M. (2003). Majority versus minority influence: When, not whether, source status instigates heuristic or systematic processing. *European Journal of Social Psychology*, 33(3), 313–330. https://doi.org/10.1002/ejsp.146

Martin, R., Martin, P. Y., Smith, J. R., & Hewstone, M. (2007). Majority versus minority influence and prediction of behavioral intentions and behavior. *Journal of Experimental Social Psychology*, 43(5), 763–771. https://doi.org/10.1016/j.jesp.2006.06.006

Meeus, W. H., & Raaijmakers, Q. A. (1986). Administrative

obedience: Carrying out orders to use psychological-administrative violence. *European Journal of Social Psychology*, 16, 311–324.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4), 371–378. https://doi.org/10.1037/h0040525

Milgram, S. (1974). *Obedience to Authority: An Experimental View*. Harper & Row.

Milgram, S., Bickman, L., & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, 13(2), 79–82. https://doi.org/10.1037/h0028070

Moscovici, S., Lage, E., & Naffrechoux, M. (1969). Influence of a consistent minority on the responses of a majority in a colour perception task. *Sociometry*, 32(4), 365–380. https://doi.org/10.2307/2786541

Mugny, G., & Papastamou, S. (1981). When rigidity does not fail: Individualization and psychologicalization as resistance to the diffusion of minority innovation. *European Journal of Social Psychology*, 10(1), 43–61. https://doi.org/10.1002/ejsp.2420100104

Mullen, B. (1983). Operationalizing the effect of the group on the individual: A self-attention perspective. *Journal of Experimental Social Psychology*, 19(4), 295–322. https://doi.org/10.1016/0022-1031(83)90025-2

Nemeth, C. J., & Kwan, J. L. (1987). Minority influence, divergent thinking, and the detection of correct solutions. *Journal of Applied Social Psychology*, 17(9), 788–799. https://doi.org/10.1111/j.1559-1816.1987.tb00339.x

Rohrer, J. H., Baron, S. H., Hoffman, E. L., & Swander, D. V. (1954). The stability of autokinetic judgments. *Journal of Abnormal and Social Psychology*, 49(4, Pt. 1), 595–597. https://doi.org/10.1037/h0060827

Sherif, M. (1936). *The psychology of social norms*. Harper & Brothers.

Sumner, W. G. (1906). *Folkways*. Ginn and Company.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport

and its nonverbal correlates. *Psychological Inquiry*, 1(4), 285–293. https://www.jstor.org/stable/1449345

Tickle-Degnen, L., & Rosenthal, R. (1992). Nonverbal aspects of therapeutic rapport. In R. S. Feldman (Ed.), *Applications of nonverbal behavioral theories and research*. Psychology Press.

Turner, J. C. (1991). *Social influence*. Thomson Brooks/Cole Publishing Co.

Wilder, D. A. (1977). Perception of groups, size of opposition, and social influence. *Journal of Experimental Social Psychology*, 13(3), 253–268. https://doi.org/10.1016/0022-1031(77)90047-6

XII. COMMUNICATING ABOUT SCIENCE

Learning Objectives

- Describe the role of science communication in modern society.
- Identify the four attributes of science literacy.
- Describe modern science commination mediums and their positive and negative impacts.

Introduction

Though my soul may set in darkness, it will rise in perfect light; I have loved the stars too truly to be fearful of the night

– Twilight Hours (1868) by Sarah Williams

Science is complicated and often tedious in its execution. However, it is also interesting, inspiring and has the potential to change people's lives. While most scientists spend their time researching in their respective fields of interest, it is also their duty to keep the public informed about recent developments and findings. This will raise public awareness and interest in science, influence their attitudes and behaviour as well as address social problems. On a larger scale, science communication also informs policy decisions made by governments that have far-reaching consequences in society.

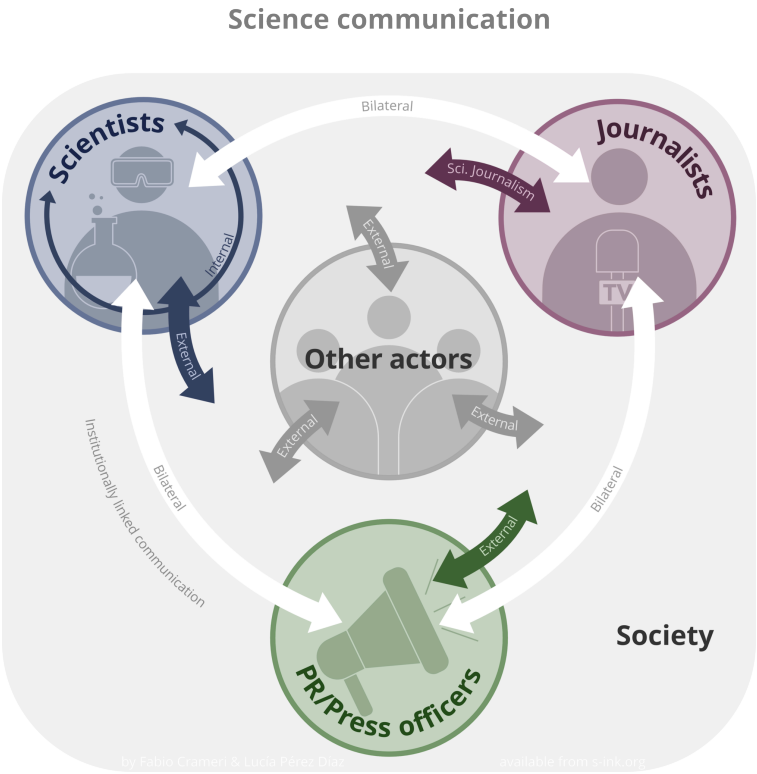


Figure 12. 1 *Schematic overview of the field and the actors of science communication according to Carsten Könneker Jessica/Feb 20: Image is missing its license*

While we have already explored how scientists share information between themselves (through peer-review), in this chapter, we will discuss how scientists share information with non-scientists. The forms of science communication can vary in form from science journalism, health communication, podcast appearances, and public workshops and exhibitions. There are also people who popularise scientific subjects through documentaries, films, and popular-science books. They can include scientists (like Richard Dawkins or Neil deGrasse Tyson) as well journalists and filmmakers (like David

Attenborough). Such communicators may focus on specific topics that are of interest to them (evolution for Richard Dawkins; language and thinking for Steven Pinker; or Astrophysics for Neil deGrasse Tyson) and help raise awareness about these topics.

Figure 12. 2 Richard Dawkins and Neil DeGrasse Tyson are popular communicators of science. In addition to their peer-reviewed publications, they have published numerous popular science books aimed at the general public. Jessica/Feb 20: Image is missing its license

Thomas and Durant (1987) state that increasing scientific literacy has numerous advantages from the economic competitiveness of having more engineers and scientists in a nation to the individual benefits of understanding critical thinking and technology. In addition, an informed electorate can promote greater democratic values in a society (Gregory & Miller, 1998).

Media Attributions

Jessica/Feb 20: Attributions might need to be adjusted

- **Figure 12.1** Fabio Crameri and Lucía Pérez Díaz https://s-ink.org/science-communication
- **Figure 12.2** Flickr and ARPAE Energy

Methods of Science Communication

"He who knows only his own side of the case knows little of that. His reasons may be good, and no one may have been able to refute them. But if he is equally unable to refute the reasons on the opposite side, if he does not so much as know what they are, he has no ground for preferring either opinion... Nor is it enough that he should hear the opinions of adversaries from his own teachers, presented as they state them, and accompanied by what they offer as refutations. He must be able to hear them from persons who actually believe them...he must know them in their most plausible and persuasive form."

— John Stuart Mill, On Liberty

The history of science communication goes back to the books written by natural philosophers which had to be coped out by hand. This meant their ideas would only be available to the very rich who could afford it. The invention of printing and the subsequent move away from Latin towards vernaculars in Europe led to greater availability of science publication to the general public. Later developments such as the radio, television and now online social media have all contributed to more avenues for reaching the public. In all cases, the issue of what is considered *the public* is also in question. Is it just middle-income people who can afford the time to spend on such matters? Is it students with some interest in going into the science? Or can it be anyone? Some promoters of scientific literacy may ridicule the public for their ignorance. On the other hand, there are also people who romanticize the *wisdom of crowds* and common sense as an alternative to scientific expertise. As Priest (2009) puts it, the role of science communication has to be to help non-scientists feel included in the scientific journey of exploring the world.

Public trust in science has its roots in public understanding of

science. Efforts to quantify such understanding often involved surveys. Bauer, Allum and Miller (2007) categorised four attributes of scientific literacy:

- knowledge of basic scientific facts
- understanding of the scientific method
- appreciation for the positive outcomes of science and technology
- rejection of superstitious beliefs, such as astrology

Bauer *et al.* (2007) focus more on attitudes towards science and technology rather than knowledge. Others (such as Durant *et al.*, 1989; and Eurobarometer opinion surveys about climate change) have been more concerned about knowledge access. We will elaborate on this topic further when we explore public understanding of science.

Inclusion and Cultural Differences

As we saw in Chapter 1, modern science has its history in natural philosophy and connects several global civilizations. However, the scientific revolution started in Western Europe followed by European imperialism and colonialism. This intrinsically linked science as a Western knowledge system which excluded colonised populations as well as women and other minority groups. Prestigious prizes in science (such as the Nobel prize) lean heavily towards Western scientists and other groups have often been systematically excluded from the scientific endeavour.

Pride and prejudice

In 2005 *Science*, on of the most prestigious peer-reviewed journals in the world, published an article about the successful cloning of 11 person-specific stem cells using 185 human eggs (Hong, 2008). Hwang Woo-suk of South Korea's Seoul National University became a national sensation. He was declared South Korea's supreme scientist while *Time* magazine listed him among the "People Who Mattered 2004" and one of the most influential people "The 2004 Time 100." Rumours circulated that Hwang may be the first to bring a Nobel prize in science to South Korea, an award that was much sorted as a source of national pride and prestige.

Later on in 2005, it came to light that Hwang had pressured subordinates who worked for him into donating their eggs. Other eggs were found to have been sourced from paid donors (a clear violation of bioethics). In addition, further investigation showed that Hwang had fabricated most of his data and results. The papers that Hwang published were retracted. Hwang was also found guilty of embezzlement and ethics violations.

Part of the reason that Hwang was able to get away with such crimes was the national sentiment about having a world-famous scientist represent them in the world stage. In addition, the low status of women without children in traditional Korean society may have led to such women thinking they were contributing to society by taking part in a (potentially) harmful procedure.

Such cases illustrate how people may manipulate the public for their own ends using science and national pride as an excuse. The great respect accorded to scientists and experts also led to people unquestioningly submitting to the demands of a man who turned out to be an unethical "scientist" and criminal.

In addition to the exclusion of some groups, others have been actively exploited by scientists. The Tuskegee Syphilis Study is a famous example of African American sharecroppers being denied life-saving syphilis medication by physicians so that they can be studied on how the disease progressed. In Canada, the residential school system indoctrinated Indigenous children into following Christianity and tried to associate scientific literacy with religion so as to stamp out Indigenous religious identity and practices. Such historical injustices have created an (understandable) suspicion of

modern science in these communities as they see it as the tool of the oppressor.

Another issue has been the promotion of non-scientific topics as alternative knowledge systems to modern science. Such endeavours often use historical injustices as an excuse to associate science with oppression and therefore, promote their own knowledge systems as the moral alternative. Fears of being called prejudiced or xenophobic can cause a chilling effect where scientists might self-censor themselves and not challenge these alternative systems. This can be counterproductive as it diminishes public appreciate for science and can even be harmful if people do not follow medical advice or government safety initiates.

However, as we have seen throughout this book, science is a human endeavour and not associated with any culture or civilization. Surveys of populations in Western societies have shown no greater understanding or appreciation of science compared to non-Western ones. In addition, anyone (regardless of gender, ethnicity, sexuality or other identity) can pursue and appreciate science as it is the only true human universal. The job of scientists is to ensure that science is viewed as such by the general population and not used as the tool of oppression and exclusion as it was used in the past.

Public Understanding of Science

One of the ways in which the masses used to get information was through traditional journalism (newspapers, magazines, radio and television). There was an expectation from the public that professional journalists would produce high quality information for general consumption. However, there was always the issue of undue influence from powerful externalities (the government or the rich) which could sway public opinion. Also, even if the information is not biased, once the information is out there, journalism is a one-way stream with no public input or dialogue. Scientists are also often frustrated with the over-simplification of their findings or dramatizing the results (Jamieson *et al.*, 2017).

Fusion and confusion

In physics there are two F-words: *fission* and *fusion*. Nuclear fission is when the nucleus of an atom splits into smaller nuclei. This process releases a large amount of energy which can be used destructively (as a nuclear bomb) or productively (as a source of nuclear energy). However, fission also produces dangerous byproducts such as toxic radiation and nuclear waste. Nuclear fusion, on the other hand, fuses to nuclei to form a larger atomic nucleus. This binding also results in the release of energy. However, this requires a lot of energy input and physicists have been trying for decades to come up with a fusion reaction that

can produce more energy than it absorbs. This is incredibly difficult as all known fusion experiments require more energy to create nuclear fusion than the energy that is released.

In 1989, two electrochemists at the University of Utah held a press conference that shook the world. Martin Fleischmann and Stanley Pons reported that they had devised an apparatus that can create nuclear fusion reaction using heavy water and palladium. The keywords in this news conference were "simple" and "excess heat." The term "Cold Fusion" was used in news media outlets as this experiment was unlike the "hot" nuclear fusion reactors that physicists were using.

However, scientists were baffled as to why a new conference was called in the first place. The two electrochemists had submitted a paper for peer-review which had been accepted. But to release the news prematurely seemed odd. Fleischmann and Pons became celebrities and toured the conference circuit discussing their experiment. Initially, other universities around the world from India to the USSR were reporting that they had replicated the results. This excitement was brief-lived as other scientists with more precise equipment were unable to replicate the results. They found that the reported excess heat was a mistaken reading that used incorrect formulas. If the paper had gone through proper peer-review, the authors might have had this pointed out to them. However, the fact that they jumped the gun and released the news to the public led to widespread criticism and the eventual downfall of both scientists. The public outcry was palpable as many people felt they had been

cheated out of what was initially publicised as "cheap, unlimited energy."

This is an example of pathological science (Change, 2004) where people are tricked into false results through wishful thinking or biases. Fleischmann and Pons were not frauds. They were genuinely convinced of their results and were well-respected in their fields before this incident. However, their initial zeal to be the first to publicise their results resulted in them having to backtrack some of their initial claims and eventually be seen as unreliable scientists.

As we can see in Figure 12.3, public opinions about scientific facts such as climate change are not really informed by facts and scientific understanding (Ballew *et al.*, 2022). Instead, political affiliation appears to be the main driving factor. If people are more prone to watch misleading opinion outlets such as Fox News, they are more likely to think of climate change as a hoax. If there are powerful corporate interests that have an incentive to keep public opinion mixed about such issues, then democratic societies will have a hard time making informed, long-term decisions.

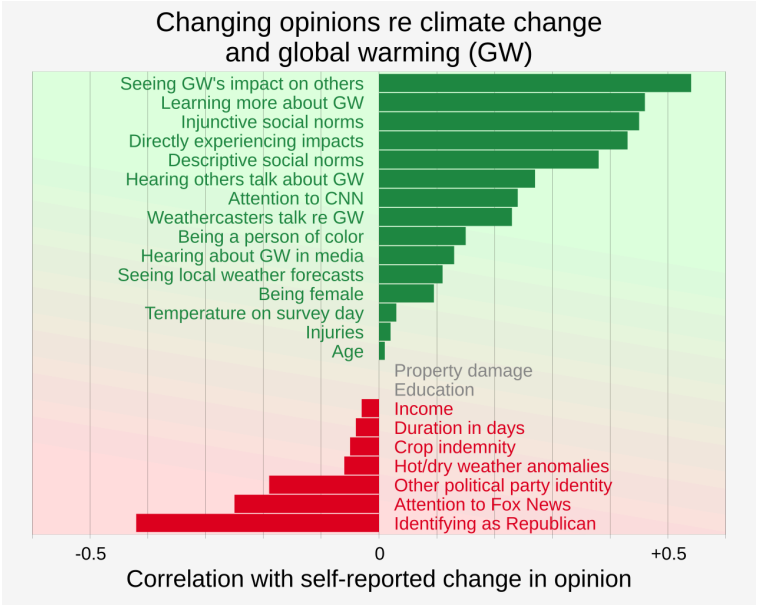


Figure 12. 3 *Correlates of self-reported changes in opinion about global warming* Jessica/Feb 20: *Image is missing its license, I am not sure if media attributions need to be adjusted*

All is not lost to traditional journalism and the undue influence of corporate interests. Websites, blogs, wikis and podcasts have taken the sceptre of public influence from traditional media in recent years. This allows for a more diverse market place of ideas and there are many scientists and science communicators who are challenging the status quo. As always, there are positive and negative consequences everything. On a positive note, the one-way communication of traditional media has changed in social media to be more open-ended. However, this also means that the social media landscape is overwhelmed with chatbots and other noise that can drown out actual science. It is an on-going challenge on how to balance free speech principles with misinformation campaigns and the influence of corporate giants who own most social media platforms.

Media Attributions

- **Figure 12.3** Ballew *et al.*, 2022 Experience with global warming is changing people's minds about it

Summary

Τὸ λοιπόν, ἀδελφοί, ὅσα ἐστὶν ἀληθῆ, ὅσα σεμνά, ὅσα δίκαια, ὅσα ἁγνά, ὅσα προσφιλῆ, ὅσα εὔφημα, εἴ τις ἀρετὴ καὶ εἴ τις ἔπαινος, ταῦτα λογίζεσθε·

Finally, brethren, whatsoever things are true, whatsoever things are honest, whatsoever things are just, whatsoever things are pure, whatsoever things are lovely, whatsoever things are of good report; if there be any virtue, and if there be any praise, think on these things.

St. Paul – Philippians 4:8

Science communication is vital for creating a vibrant, informed electorate. However, in this chapter we saw some of the challenges that scientists face in communicating effectively with the public in an environment of misinformation and political biases. This book was an attempt to address some of the issues involved in critical thinking in science. While it is far from exhaustive, I hope you have found it thought-provoking and that it started you on a journey of exploration into some fascinating topics about science.

Key Takeaways

- Scientists have a duty to keep the public informed about recent developments and findings.
- The forms of science communication can include science journalism, health communication, podcast appearances, and public workshops and exhibitions.
- The invention of movable type printing in Europe led to greater availability of science publication to the

general public.

- Four attributes of scientific literacy include knowledge of basic scientific facts, understanding of the scientific method, appreciation for the positive outcomes of science and technology, and rejection of superstitious beliefs.
- Alternative knowledge systems to modern science often uses historical injustices as an excuse to associate science with oppression.
- Science is often (incorrectly) seen as a Western knowledge system which excluded colonised populations as well as women and other minority groups when, in fact, it is a universal endeavour that anyone can pursue.
- Science communication is vital for creating a vibrant, informed society.

Jessica/Feb 20: Acknowledgements section missing.

References

Ballew, M., Marlon, J., Goldberg, M., Maibach, E., et al. (4 August 2022). Changing minds about global warming: vicarious experience predicts self-reported opinion change in the USA. *Climatic Change*. 173 (19).

Bauer, M., Allum, N. & Miller, S. (2007). What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda, *Public Understanding of Science*, 16, 80–81.

Chang, K. (2004). "*US will give cold fusion a second look*", The New York Times.

Durant, J. R., Geoffrey E. A. & Geoffrey, T. P. (1989). The public understanding of science. *Nature*. 340 (6228), 11–14.

Gregory, J. & Miller, S. (1998). *Science in Public: Communication, Culture, and Credibility*. New York: Plenum Trade.

Hong, Sungook (2008-03-01). The Hwang Scandal That "Shook the World of Science. *East Asian Science, Technology and Society*. 2 (1), 1–7.

Jamieson, K. H. Kahan, D. M. & Scheufele, D. A. (2017). *The Oxford handbook of the science of science communication*. New York, NY: Oxford University Press.

Priest, S. H. (2009). Reinterpreting the audiences for media messages about science, in Holliman *et al.* (eds), *Investigating Science Communication in the Information Age: Implications for Public Engagement and Popular Media*, Oxford: Oxford University Press, 223–236.

Thomas, G. & Durant, J. (1987). Why should we promote the public understanding of science? *Scientific Literacy Papers: A Journal of Research in Science, Education and the Public*. 1, 1–14.

This is where you can add appendices or other back matter.

Version History

This page provides a record of changes made to this learning resource, Critical Thinking. Each update increases the version number by 0.1. The most recent version is reflected in the exported files for this resource.

If you identify an error in this resource, please report it using the TRU Open Education Resource Error Form.

Version	Date	Change	Details
1.0	May 2025		Critical Thinking is published, made available for the public.
1.1	August 13, 2026	Exports generated	Exports generated for PDF, Print PDF and EPUB and made available on the book's landing page.