CRITICAL THINKING IN PSYCHOLOGY

DISPOSITIONS, COGNITIVE INSIGHTS, AND RESEARCH SKILLS

EDITORS
MARC CHAO
MUHAMAD ALIF BIN IBRAHIM

Critical Thinking in Psychology: Dispositions, Cognitive Insights, and Research Skills

CRITICAL THINKING IN PSYCHOLOGY: DISPOSITIONS, COGNITIVE INSIGHTS, AND RESEARCH SKILLS

EDITED BY MARC CHAO AND MUHAMAD ALIF BIN IBRAHIM

Critical Thinking in Psychology: Dispositions, Cognitive Insights, and Research Skills is an adaptation of the following sources:

- *How to Think Critically* by Stephanie Gibbons & Justine Kingsbury, University of Waikato, is used under a CC BY-NC Licence.
- *Neuroscience, Psychology and Conflict Management* by Judith Rafferty, James Cook University, is used under a CC BY-NC Licence.
- *Mastering Thinking: Reasoning, Psychology, and Scientific Methods* by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA Licence.
- *Research Methods in Psychology* by Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler & Dana C. Leighton, Kwantlen Polytechnic University, is used under a CC BY-NC-SA Licence.

First published in 2025 by James Cook University

How to cite this book:

Chao, M., & Bin Ibrahim, M. A. (Eds.). (2025). *Critical thinking in psychology: Dispositions, cognitive insights, and research skills*. James Cook University. https://doi.org/10.25120/5a9n-kegr

CONTENTS

Chapter 7: Research Ethics

Chapter 8: Psychological Measures

Chapter 9: Research Methods in Psychology

Chapter 10: Descriptive Statistics

Chapter 11: Inferential Statistics

ACKNOWLEDGEMENT OF COUNTRY

At James Cook University we acknowledge with respect the Aboriginal and Torres Strait Islander peoples as the first peoples, educators and innovators of this country. We acknowledge that Country was never ceded, and value the accumulation of knowledge and traditions that reflect the wisdom of ancestral lines going back some 60,000 years, and recognise the significance of this in the ways that Aboriginal and Torres Strait Islander peoples are custodians of Country. As a University, we will continue to learn ways to care for and be responsible for Country, and we will collectively seek to build a future that is based on truth-telling, mutual understanding, hope, empowerment, and self-determination.

Kassandra Savage (JCU Alumni), 'Coming Together and Respecting Difference', acrylic on canvas, 2014, 90cm x 90cm. © Kassandra Savage, reproduced with permission of the artist.

ABOUT THE EDITORS

Dr Marc Chao

Marc is a Lecturer in Psychology with the College of Healthcare Sciences at James Cook University Bebegu Yumba (Townsville) Campus, where he also serves as the Academic Advisor for the Bachelor of Psychological Science.

His research applies psychological science to public health challenges, focusing on mindfulness-based interventions, global consciousness, critical thinking to counter misinformation, and associative learning in human behaviour. By integrating these areas, Marc seeks to develop culturally responsive, evidence-based strategies that strengthen mental health, resilience, and adaptive behaviour, particularly in vulnerable and underserved populations. His doctoral research examined learning across multiple contexts in classical and instrumental conditioning and its influence on the recovery of first-learned associations, offering new insights into extinction learning mechanisms.

Marc is a passionate educator whose philosophy centres on equitable access to learning and inclusive teaching practices. He has consistently achieved high subject (4.9/5 for SP52 2023 PY2111) and teacher ratings (5/5 for SP52 2023 PY2111) and is a Fellow of the Higher Education Academy (FHEA).

His qualifications include a PhD in Psychology from James Cook University (passed with no amendments), an MSc in Psychology and Neuroscience of Mental Health (Distinction) from King's College London, an MSc in International Business Management (Top Graduate) from Edinburgh Napier University, a Post-Graduate Certificate in Academic Practice (Distinction) from the University of Liverpool, and a Specialist Diploma in Career Counselling from Republic Polytechnic, Singapore (endorsed by the Career Industry Council of Australia).

Dr Muhamad Alif Bin Ibrahim

Alif (pronouns: he/him) has a PhD in Society and Culture from James Cook University, Australia. His doctoral study examined the resilience of LGBTQ couples against the socio-political stressors they experienced in Singapore. His research provided pertinent insights into the challenges that LGBTQ couples face in Singapore and the ways in which they cope and maintain their relationships in a socio-political context that continues to marginalise their relationships and families.

As a social scientist, Alif draws from the disciplines of psychology and sociology. His research interests include relationships and families in the LGBTQ community, qualitative research methodologies, coping and resilience, and critical approaches to health and illness. Alif is also actively involved with various internationally recognised societies, including being an International Representative of the International Society for Critical Health Psychology (ISCHP) and a New Scholar Member under the Publications Committee of the International Association for Relationship Research (IARR).

Beyond his PhD research, he has over 12 years of research experience utilising social science knowledge, theories, and methods at the intersections of health and public health. He is experienced in utilising such knowledge and tools to develop and improve health services and interventions for healthcare workers and individuals from the LGBTQ community. Alif is also a passionate advocate for social justice and equality. He believes that research can be a powerful tool for promoting social change.

INTRODUCTION

By Marc Chao

The Need for Critical Thinking in Psychology

In an era of rapid information exchange, misinformation, disinformation, competing narratives, and complex societal challenges, the ability to think critically has never been more crucial. In psychology, critical thinking is not just an academic skill; it is a fundamental tool for understanding human behaviour, evaluating research, and making informed decisions. Whether in clinical practice, experimental research, or everyday reasoning, the ability to assess evidence, question assumptions, and avoid cognitive biases is essential.

This book, *Critical Thinking in Psychology: Dispositions, Cognitive Insights, and Research Skills*, equips readers with key dispositions, cognitive awareness, and research-driven analytical skills essential for rigorous thought. It integrates these elements to foster critical engagement with information, enabling readers to assess evidence, recognise biases, and apply structured methods for evaluating psychological research and claims.

A Three-Factor Approach: Dispositions, Cognitive Insights, and Research Skills

Critical thinking in psychology requires a multifaceted approach that goes beyond intuition and general reasoning. This book emphasises a three-factor approach to critical thinking by integrating dispositions, cognitive insights, and research skills. Dispositions refer to the essential traits that cultivate a critical thinking mindset, including scepticism, intellectual honesty, curiosity, and open-mindedness. These qualities influence how individuals approach information, assess claims, and engage with diverse perspectives, shaping their ability to think critically and objectively.

Cognitive insight and reasoning errors focus on the ways biases, logical fallacies, and flawed thinking distort our ability to assess information accurately. Recognising these mental shortcuts and errors allows individuals to safeguard against misinformation and flawed interpretations of data, helping them navigate the complexities of decision-making in psychology and beyond.

Research and analytical skills form the third pillar of this approach, ensuring that critical thinking is applied in a structured and systematic manner. The book provides guidance on evaluating experimental designs, understanding statistical reasoning, and critically appraising psychological studies. It introduces frameworks such as the OBSERVE method, offering a systematic approach to inquiry and evidence evaluation, ultimately enhancing the ability to analyse research and draw well-founded conclusions.

These three topics: dispositions, cognitive insights, and research skills are interconnected and build upon each other throughout the book. Dispositions form the foundation by shaping the mindset with which students approach information and learning. Cognitive insights then highlight the common reasoning errors and biases that can undermine even well-intentioned thinking. Finally, research skills provide the practical tools and structured methods for applying critical thinking in psychological inquiry. While students may choose to focus on sections most relevant to their immediate needs, such as improving research skills, the book is designed to be read as a cohesive whole. Each part enriches the others, offering a layered understanding of how to think critically in psychology. To gain the most from this resource, readers are encouraged to first develop an awareness of their own dispositions, then sharpen their ability to recognise cognitive pitfalls, and finally apply these insights through the research and analytical techniques introduced later in the book.

By incorporating these three aspects, this book offers a comprehensive framework for developing critical thinking skills that are directly applicable to psychological research, academic study, and everyday decision-making.

What to Expect from This Book

Throughout this book, we explore core components of critical thinking in psychology, breaking them down into practical, engaging discussions. The book is structured into the following key sections:

- **Foundations of critical thinking:** Understanding what critical thinking is, how it differs from common sense, and why it is indispensable in contemporary society.
- **The OBSERVE framework:** Applying a structured approach to inquiry to ensure that reasoning is evidence-based and free from misleading influences.
- **Arguments and reasoning:** Learning how to construct and analyse arguments, distinguish between deductive and inductive reasoning, and evaluate arguments for soundness, cogency, and strength.
- **Logical and informal fallacies:** Recognising fallacies in arguments and understanding how faulty reasoning can lead to misinformation.
- **Cognitive biases and beliefs:** Examining how perception, memory, attention, and belief formation influence our reasoning and how biases distort thinking.
- **Knowledge and science:** Exploring how knowledge is acquired, the methods of science, and how scientific reasoning differs from common sense.

- **Research methods and ethics:** Understanding the scientific method, research ethics, psychological measurement, and the various research designs used in psychology.
- **Research and analytical skills:** Developing skills to analyse psychological studies, interpret data, and apply statistical reasoning.

Each chapter provides theoretical insights and real-world applications to reinforce critical thinking skills in psychology. Whether you are a student, researcher, or professional, this book will serve as a valuable resource for developing a sharper, more discerning approach to evaluating information and conducting research.

A Call to Active Engagement

Critical thinking is not a passive endeavour. It is an active, deliberate process that requires continuous reflection and practice. As you progress through this book, you are encouraged to engage with the material actively: question assumptions, apply concepts to real-life situations, and critically assess your own thought processes. By doing so, you will not only refine your cognitive skills but also cultivate a mindset that values reasoned inquiry and intellectual rigour.

In psychology, as in life, the ability to think critically is a powerful asset. It enables us to navigate the complexities of human behaviour, challenge misleading claims, and contribute to a more informed and rational discourse. Let this book be your guide in mastering the art of critical thinking, both as a discipline and as a way of life.

CHAPTER 1: FOUNDATIONS OF CRITICAL THINKING

In a world saturated with information, opinions, and competing narratives, critical thinking stands as an essential skill for navigating complexity, making informed decisions, and fostering intellectual growth. At its core, critical thinking is the disciplined process of actively analysing, evaluating, and synthesising information to arrive at well-reasoned conclusions. It transcends mere problem-solving or decision-making, offering a structured approach to understanding and addressing multifaceted issues with clarity and precision.

This chapter delves into the principles, processes, and practices of critical thinking, exploring how biases, emotions, and assumptions can cloud our judgement and hinder objective analysis. From understanding the difference between common sense and critical reasoning to recognising the role of creativity in problem-solving, this chapter highlights how critical thinking is not merely an abstract academic exercise but a practical tool applicable across all aspects of life.

We will also examine the interplay between creativity and critical thought, demonstrating how both are interdependent in fostering innovation and insight. Additionally, the chapter introduces key dispositions and skills essential for critical thinking, such as intellectual honesty, scepticism, and open-mindedness, alongside practical frameworks like the OBSERVE method to guide structured inquiry and analysis.

By mastering the principles and tools of critical thinking, individuals can become more reflective, adaptable, and informed, capable of discerning truth from misinformation and approaching challenges with both scepticism and curiosity. As we embark on this exploration, we invite you to question assumptions, embrace alternative perspectives, and engage with the material actively, laying the groundwork for a more thoughtful, insightful, and reasoned approach to understanding the world.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand critical thinking:** Define critical thinking and explain its significance in personal, societal, and workplace contexts.
- **Differentiate between critical thinking and common sense:** Identify the key

differences between critical thinking and common sense, including their reliance on systematic evaluation versus intuition.

- **Recognise the role of bias:** Understand the impact of cognitive biases, such as confirmation bias, on judgement and decision-making, and explore strategies to mitigate their influence.
- **Appreciate the importance of critical thinking today:** Explain why critical thinking is essential in the information age, particularly for navigating misinformation and the "attention economy".
- **Link creativity with critical thinking:** Describe how creativity and critical thinking complement each other in generating and evaluating ideas and solutions.
- **Adopt critical thinking dispositions:** Cultivate dispositions like curiosity, intellectual honesty, scepticism, and open-mindedness to support a critical thinking mindset.
- **Develop critical thinking skills:** Apply key skills such as analysing arguments, understanding statistical concepts, and evaluating experimental designs in various contexts.
- **Apply the OBSERVE framework:** Use the OBSERVE framework to systematically assess phenomena, develop hypotheses, evaluate evidence, and draw conclusions based on the criteria of adequacy.

1.1. CRITICAL THINKING

By Marc Chao, adapted from Michael Ireland

Critical thinking is the disciplined process of actively and skilfully conceptualising, applying, analysing, synthesising, and evaluating information gathered from observation, experience, reflection, reasoning, or communication. It is a cornerstone of psychological inquiry and practice, enabling us to navigate the complexities of human behaviour and mental processes with clarity and precision.

At its core, critical thinking involves questioning assumptions, evaluating evidence, and considering alternative perspectives. This process is not merely an academic exercise but a vital skill that empowers individuals to make informed decisions, solve problems effectively, and understand the world more deeply.

One of the most challenging aspects of critical thinking is overcoming our inherent biases. Bias refers to a systematic deviation from rationality in judgement or decision-making. It occurs when individuals' perceptions and interpretations are influenced by their pre-existing beliefs, preferences, or experiences, leading to skewed or partial viewpoints. In psychology, biases can manifest in various forms, such as cognitive biases, which affect how we process information, and affective biases, which are influenced by our emotions. Recognising and mitigating biases is crucial for critical thinking, as it allows for more objective and balanced analysis of information.

Humans are naturally inclined to accept information that aligns with their pre-existing beliefs and to reject information that contradicts them. This cognitive bias, known as confirmation bias, can significantly hinder our ability to think critically. When we encounter information that supports our beliefs, we tend to accept it without much scrutiny. Conversely, when we face information that challenges our beliefs, we often become defensive and dismissive.

This defensiveness is a natural human response, often rooted in our desire to protect our sense of identity and coherence. Our beliefs are not just abstract ideas; they are integral to how we understand ourselves and the world around us. When these beliefs are questioned, it can feel like a personal attack, triggering an emotional reaction. Emotions like anger, frustration, and fear can take over, making it difficult to engage with the new information objectively. Instead of considering the evidence on its merits, we might focus on discrediting the source or finding flaws in the argument. This emotional reaction can create a barrier to understanding, preventing us from seeing the potential value in the information being presented.

For example, consider a person who strongly believes in a particular political ideology. When presented with evidence that supports their views, they are likely to accept it readily. However, when confronted with evidence that contradicts their beliefs, they may question the credibility of the source or the validity of the data. This defensiveness is a natural response, rooted in our desire to maintain a consistent and coherent worldview.

Moreover, when we are defensive, we are more likely to shut out information that might be useful. This is particularly problematic in an age where misinformation and disinformation are rampant, because defensiveness does not just filter out poor-quality information. Instead, it can lead us to dismiss high-quality evidence simply because it challenges our existing beliefs. At a time when reliable information can be harder to identify, overlooking such evidence is especially harmful to our understanding and decision-making. Misinformation refers to false or inaccurate information shared without the intent to deceive, while disinformation is deliberately false information spread with the intent to mislead. By dismissing information that contradicts our beliefs, we may miss out on important insights and opportunities for growth. This tendency to reject challenging information can reinforce our existing biases and limit our ability to learn and adapt.

Psychologist Daniel Kahneman offers a helpful framework to understand why these biases are so pervasive. In his book Thinking, Fast and Slow, he describes two modes of thinking: System 1 and System 2. System 1 is fast, intuitive, and automatic. It enables quick judgements but is also prone to errors and biases. System 2 is slow, deliberate, and analytical, requiring effort to engage. Most of our daily thinking relies on System 1, which is efficient but often irrational. Critical thinking requires us to consciously override our automatic System 1 responses and activate System 2, enabling us to evaluate evidence more carefully and make more reasoned decisions.

Therefore, critical thinking requires us to recognise these biases and actively work to mitigate their influence. It involves being open to new information, even when it is uncomfortable or challenging. This openness is not about abandoning our beliefs but about being willing to re-evaluate them in light of new evidence. It is about fostering a mindset of curiosity and scepticism, where we are constantly questioning and refining our understanding of the world.

Moreover, critical thinking is not a solitary endeavour. It thrives in environments where diverse perspectives are valued and where dialogue and debate are encouraged. Engaging with others who hold different viewpoints can help us to see our own biases more clearly and to develop a more nuanced understanding of complex issues. In this way, critical thinking is both a personal and a collective practice, essential for the advancement of knowledge and the betterment of society.

In the field of psychology, critical thinking is particularly crucial. Psychologists must navigate a vast array of theories, research findings, and clinical practices, each with its own set of assumptions and evidence. By applying critical thinking, psychologists can discern which theories and practices are most supported by evidence, which are most effective in different contexts, and how best to integrate new findings into their work.

As we delve deeper into the principles and practices of critical thinking in psychology, we will explore various strategies and tools that can help us to think more critically. From understanding logical fallacies to developing better questioning techniques, this journey will equip us with the skills needed to navigate the complexities of human thought and behaviour with greater insight and effectiveness.

Here is a brief video by Macat that explains the concept of critical thinking [2:30]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=791#oembed-1

Critical Thinking and Common Sense

Critical thinking is often misconstrued as common sense, leading many to dismiss it as unnecessary to learn or develop. This misconception arises from the belief that common sense, which is generally understood as sound practical judgement, is sufficient for navigating complex issues and making decisions. However, critical thinking and common sense are fundamentally different in several important ways.

Common sense is based on everyday experiences and intuitive judgements that seem obvious to most people. While it can be useful in routine situations, common sense is inherently influenced by personal biases, cultural norms, and subjective perceptions. These factors can cloud objective thinking and lead to errors in judgement. For example, common sense might suggest that a correlation (correlation is also known as a relationship) between two events implies causation, but critical thinking teaches us to scrutinise such assumptions and consider alternative explanations.

Critical thinking, on the other hand, is a disciplined and systematic approach to evaluating information and arguments. It involves skills such as interpreting, analysing, evaluating, and inferring, which go beyond the surface-level understanding that common sense provides. Critical thinking requires us to question assumptions, seek out evidence, and consider multiple perspectives before drawing conclusions. This rigorous process helps to mitigate the influence of biases and ensures that our judgements are based on sound reasoning and reliable evidence.

Moreover, common sense often relies on heuristics, or mental shortcuts, that can lead to cognitive biases. These biases, such as confirmation bias and availability heuristic, can distort our perception of reality and hinder our ability to think critically. For instance, confirmation bias leads us to favour information that confirms our pre-existing beliefs and ignore information that contradicts them. For example, if someone believes that left-handed people are more creative, they might pay more attention to instances where left-handed individuals display creativity and overlook instances where right-handed individuals do the same.

Similarly, the availability heuristic causes us to overestimate the likelihood of events based on how easily examples come to mind. For instance, after seeing news reports about aeroplane accidents, a person might overestimate the danger of flying, despite statistics showing that air travel is safer than car travel. Critical thinking, by contrast, encourages us to be aware of these biases and actively work to counteract them.

Another key difference is that common sense is typically reactive, responding to situations as they arise, whereas critical thinking is proactive. Critical thinkers anticipate potential problems, gather relevant information, and plan their actions accordingly. For example, a person relying on common sense might assume that a "low-fat" processed food is healthy because the label suggests it, whereas a critical thinker would read the nutritional information, investigate ingredients, and make decisions based on evidence rather than marketing claims. This proactive approach enables them to make more informed and effective decisions, even in complex and uncertain situations. Similarly, in a work setting, a common-sense approach might involve dealing with conflicts as they occur, while a critical thinker would establish clear communication channels and conflict resolution strategies in advance to handle potential disputes more effectively.

Hence, while common sense can be helpful in everyday situations, it is not a substitute for critical thinking. Critical thinking provides a more robust and reliable framework for evaluating information and making decisions. By recognising the limitations of common sense and embracing the principles of critical thinking, we can enhance our ability to navigate the complexities of the modern world with greater clarity and insight.

Critical Thinking in Contemporary Society

In today's fast-paced, media-saturated, and increasingly politicised world, critical thinking skills are more essential than ever. We are inundated with an overwhelming amount of information, news, opinions, and ideas at an unprecedented speed. This deluge can feel inescapable, making it crucial to develop sharp, refined thinking skills to navigate this environment effectively. The rapid changes in our information landscape are set to continue accelerating, presenting us with more information than we can possibly absorb and process. Consequently, we must make daily decisions about which sources to trust and which to ignore. It is vital to avoid confining ourselves to sources that only reinforce our pre-existing beliefs, such as worldviews and political beliefs.

Historically, this is a relatively new challenge. In the span of about 80 years, the global population has transitioned from being largely illiterate to becoming active users of the internet, the most extensive information source ever created. This dramatic shift means that while the majority of adults today are active internet users, we are also living in what has been termed a 'post-truth era'.

The consequences of this post-truth era are severe. We face numerous social, economic, and environmental issues exacerbated by a lack of critical thinking. For instance, unvaccinated populations contribute to the resurgence of previously eradicated diseases, and misinformation spreads so widely that NASA has had to debunk absurd claims like Mars being a secret child labour colony. Similarly, during the COVID-19 pandemic, misinformation about the virus and vaccines spread widely on social media, leading to vaccine hesitancy and resistance. This misinformation had severe public health consequences, including increased infection rates and deaths. In another example, the August 2024 riots in the UK were fuelled by misinformation following a tragic incident in Southport, where false claims about the identity of the attacker spread rapidly online. This misinformation incited violence and unrest, particularly targeting mosques and asylum-seeker accommodations.

We are bombarded with a constant stream of information and ideas, making it more urgent than ever to practice skilled and active interpretation and evaluation of observations, communications, information, and arguments. In simpler terms, critical thinking helps us filter useful signals from the overwhelming noise. Without it, we cannot reliably determine the accuracy and usefulness of the information we encounter.

While this challenge is not entirely new, it has become more pressing as the volume of information and rhetoric grows daily. In the twenty-first century, your attention is a valuable commodity, with countless corporations and interests competing for your engagement. This 'attention economy' has led media companies to employ cognitive scientists, social scientists, and statisticians to design platforms that exploit human vulnerabilities, primarily through emotional manipulation, to keep you engaged. This is addiction science weaponised to turn your attention into a commodity for sale. Be mindful that the information you consume is often heavily processed, much like processed food, and can be just as harmful to your mental health.

In the broader commercial context, information is a major commodity in the twenty-first-century global economy, hence the term 'Information Age'. We are entering the 'Fourth Industrial Revolution', characterised by information and information technologies. As a result, employers increasingly recognise the importance of critical thinking among their employees. Many businesses now have dedicated manuals and guides to adopt critical thinking approaches, and it is considered one of the primary skill sets for success in various industries.

Recently, global employers ranked the top ten skills, with half directly related to critical thinking. The top three skills—complex problem-solving, critical thinking, and creativity—are core components of critical thinking. Other critical thinking-related skills include judgement and decision-making, and cognitive flexibility.

Here are the top 10 skills identified by the World Economic Forum:

1. Complex problem-solving
2. Critical thinking
3. Creativity
4. People management
5. Coordinating with others

6. Emotional intelligence
7. Judgement and decision-making
8. Service orientation
9. Negotiation
10. Cognitive flexibility.

The question 'Who should study critical thinking?' closely intertwines with 'Why study critical thinking?'. The benefits of studying critical thinking are numerous and there are genuinely no downsides to developing a critical thinking skillset.

Here is a video by Bart Millar that explains why critical thinking is important [15:24]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=791#oembed-2

References

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under CC BY-SA licence.

1.2. CREATIVITY AND CRITICAL THINKING

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Creativity versus critical thinking is a fascinating topic that often goes unexamined. You might wonder, "What does creativity have to do with critical thinking?". However, these two concepts are closely linked and even interdependent.

Robert E. Franken (1994, p. 396) offers a popular definition of creativity, emphasising that the acts of creating and recognising ideas, alternatives, and possibilities are inherently creative. This definition highlights that recognising novelty is as much a part of creativity as generating it. Creativity is crucial in problem-solving, communication, and entertainment. While we often associate creativity with producing entertainment (such as music, film, art, and dance), it is equally vital in our daily activities of communication and problem-solving.

Another insightful definition comes from Ghuman and Aswathappa (2010, p. 540), who explain that creativity involves generating new ideas, challenging assumptions, and viewing things from alternate perspectives. These activities are integral to critical thinking. Many people may not realise that challenging assumptions, seeing things from different viewpoints, and developing new ideas are creative acts.

From these definitions, it is clear that creativity extends beyond painting a picture or writing a song. It is a commonplace, day-to-day activity. Creativity is not a mystical power possessed by a few geniuses; it is a skill we all have to varying degrees. Our level of creativity depends on our interests, experiences, and training. Without creativity, we would be immobilised in life, unable to overcome the first hurdle we encounter.

The misconception that critical and creative thinking are unrelated or even incompatible stems from outdated stereotypes. The myth of 'right brain versus left brain' suggests that the left hemisphere is for analytical, rational functions, while the right is for creativity. Some believe individuals are exclusively 'left-brained' or 'right-brained'. These claims are unfounded. Both hemispheres work together to produce critical and creative thinking, and we all use both sides of our brains.

Creativity is indispensable to many steps in critical thinking. The ability to think imaginatively about a situation, come up with new ideas, hypotheses, perspectives, and insights is essential to critical thinking. Creativity allows us to see problems in a new light and generate new solutions or use old solutions in new ways. Good critical thinking depends on mental flexibility and innovation, which are hallmarks of creativity. Critical thinking is a goal-directed activity used to achieve specific outcomes, such as interpretations, evaluations, decisions, explanations, actions, or problem-solving. Success in these areas depends on how creatively we can generate a range of options.

Consider a songwriter in a band arranging guitar riffs. As they choose notes, perfect the tempo, rhythm, and structure, they must constantly appraise their work, critically evaluating each step. This process of analysis, reasoning, and problem-solving is active critical thinking.

Similarly, imagine an investigative journalist uncovering political corruption. They gather information, evaluate its credibility and usefulness, and generate alternative explanations for events. They use imagination to view information and people from different perspectives, innovate in their investigative approaches, and find hidden patterns. This process relies heavily on creativity to be successful.

Now, consider a researcher designing a psychology experiment to study the effects of sleep deprivation on cognitive performance. The researcher must first develop a hypothesis and design an experiment that accurately tests this hypothesis. This involves selecting appropriate tasks to measure cognitive performance, determining the duration and conditions of sleep deprivation, and ensuring that the experiment controls for external factors (i.e., confounding variables) that can influence the experimental outcomes.

As the researcher plans the experiment, they must think creatively to devise innovative methods for measuring cognitive performance. For instance, they might design a new cognitive task that is both engaging and challenging, ensuring it effectively captures the nuances of cognitive decline due to sleep deprivation. They also need to consider ethical implications and find creative solutions to minimise any potential harm to participants.

During the experiment, the researcher must critically evaluate the data collected, looking for patterns and anomalies. They might notice that participants perform differently on cognitive tasks depending on the time of day, prompting them to consider additional variables such as circadian rhythms. This requires creative thinking to generate new hypotheses and design follow-up experiments to explore these observations further.

In this way, the process of designing and conducting a psychology experiment relies heavily on both creative and critical thinking. Creativity is essential for developing innovative experimental designs and solutions, while critical thinking is crucial for analysing data, identifying patterns, and drawing valid conclusions.

Hence, critical thinking and creativity are highly interdependent and similar skill sets. Creativity enhances critical thinking, and critical thinking enhances creativity. Both can be learned and developed through self-reflection and practice. Like a weightlifter building muscles, repetition and dedication are key.

Enhancing Creative Thinking

Enhancing creative thinking begins with heightened awareness and exposure. It involves paying close attention to how you think, create, tackle problems, generate ideas, and expose yourself to new concepts. A monotonous environment stifles creativity. Reflect on whether you live in an echo chamber of ideas and input. An echo chamber is an environment where a person only encounters information or opinions that reflect and reinforce their own. For example, if you only follow social media accounts that share your views, you might miss out on diverse perspectives and new ideas. Creativity thrives on new and challenging ideas, viewpoints, and perspectives. If you only echo the opinions of your favourite thought leaders, you are not an independent thinker. Instead, your voice and thinking may have been co-opted by others. Evaluate whether your opinions align entirely with any major political or ideological stance. If so, you might be ideologically captivated. For example, if knowing your stance on gun control allows me to predict your views on abortion or immigration, you may not be as independent a thinker as you believe.

To foster creativity, expose yourself to unfamiliar or even disagreeable input. Genuinely try to understand and inhabit the worlds of new and different people, including those you dislike or disagree with, and even fictional characters. Notice the limits you place on yourself and your influences. Read, listen, watch, and converse widely with others outside your usual genres, and expose yourself to diverse viewpoints. Challenge yourself to regularly try something new. It does not have to be extreme like skydiving; it could be as simple as driving down an unfamiliar road or cooking a new dish.

After introspection and raising your awareness, scrutinise your 'environment of ideas' and start practising the art of creation. This step involves rehearsing the generation of new ideas, perspectives, and solutions in a playful manner. Apply this to anything:

- How many uses of a kitchen item can you think of?
- How many activities can you list that you would like to engage in if time and money were not an issue?
- How many solutions can you come up with for a fictitious problem?

This method helps release creativity from the confines of conventional thinking. The goal is to practice loosening up your thinking and learning about yourself. Aim to produce long lists, not necessarily good ones – prioritise diverse quantity over quality. During this practice, notice the automatic tendency to evaluate and belittle certain ideas. You might find yourself thinking, 'That's dumb!', 'That's impractical!', 'That won't work!', or 'That's too similar to other options!'. Recognising this internal monologue is crucial. It is natural and part of how we evolved to survive by making instantaneous evaluative decisions. Everyone has an internal 'voice of criticism' that provides negative commentary, hindering creative pursuits. This exercise aims to minimise the influence of this critical voice during the creation stage. Recognising and relaxing this knee-jerk criticism is essential for becoming more creative. Suspend all judgement and be outlandish in your ideas. Evaluation can come later, but there is no place for condemnation when generating ideas. Evaluation throttles creativity. Repetition is key. The more you practice generating options, ideas, pathways, uses, activities, and solutions, the better a critical thinker you will become.

References

Franken, R. (1994). *Human motivation* (3rd ed.). Brooks/Cole Publishing Co.

Ghuman, K., & Aswathappa, K. (2010). *Management: Concept, practice and cases*. Tata McGraw Hill.

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

1.3. CRITICAL THINKING DISPOSITIONS AND SKILLS

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Critical thinking has been defined as purposeful reflecting and reasoning about what to do or believe when confronting complex issues, taking into account relevant context. It involves a diligent process of scrutinising beliefs or claimed knowledge against the backdrop of supporting evidence and potential outcomes.

To think critically, it is essential to adopt critical thinking skills and dispositions. Dispositions are the inherent qualities or tendencies that shape how we think and act. In the context of critical thinking, dispositions such as subject matter knowledge, metacognition (thinking about thinking), curiosity, intellectual honesty, and open-mindedness are crucial. These traits influence our approach to information and ideas, guiding us to engage thoughtfully and critically. As you read this chapter, consider how these dispositions might apply to your own thinking. Reflect on whether you are open to revising your beliefs when presented with strong evidence, whether you actively seek out perspectives that challenge your views, and whether you regularly think about how you make decisions. This brief reflection can help you become more aware of your strengths and areas where you might develop your critical thinking dispositions further.

Critical thinking encompasses several key skills or abilities, such as recognising problems, gathering and marshalling pertinent information, and interpreting data to appraise evidence and evaluate arguments. It also involves recognising unstated assumptions and values, comprehending and using language with accuracy, and drawing warranted conclusions and generalisations. By putting these conclusions to the test and reconstructing our patterns of beliefs based on wider experience, we can render accurate judgements about specific things and qualities in everyday life.

Critical Thinking Dispositions

A solid understanding of the subject matter provides the foundation for critical thinking, while dispositions such as metacognition, curiosity, intellectual honesty, and open-mindedness shape how we engage with that knowledge. These dispositions are essential psychological traits that influence how we approach and engage with information and ideas.

Subject Matter Knowledge

To think critically, it is essential to have a solid grasp of the subject matter. However, this alone is not enough. Equally important is our ability to be self-aware and introspective, a process known as metacognition. This involves scrutinising newly acquired information and reflecting on our understanding of it. The effectiveness of our critical thinking is closely tied to our foundational knowledge and our willingness to engage deeply with that knowledge.

For instance, a student might possess extensive knowledge about a particular topic but may have acquired this knowledge passively, without any reflection or contemplation. In such cases, critical thinking has not been activated. True critical thinking requires an active engagement with the material, questioning assumptions, and considering the implications of what has been learned.

Active involvement in learning processes significantly enhances critical thinking skills. When students participate actively in class through activities such as peer reviews, classroom presentations, small group discussions, film reviews, or laboratory exercises, they are more likely to develop and refine their critical thinking abilities. These interactive and participatory methods encourage students to think deeply about the subject matter, challenge their own and others' assumptions, and apply their knowledge in practical, meaningful ways.

Metacognition

Metacognition is the awareness and regulation of one's own thinking processes. It involves a deep understanding of what you know, what you do not know, and how you can improve your thinking. This self-awareness allows you to monitor, evaluate, and adjust your reasoning strategies effectively.

By engaging in metacognition, you can identify your strengths and weaknesses, set meaningful goals, plan your actions thoughtfully, check your progress regularly, and reflect on your outcomes. This reflective practice not only enhances your critical thinking skills but also fosters a mindset of continuous improvement.

Metacognition also plays a vital role in helping you become more open-minded, curious, and flexible in your thinking. It enables you to recognise your assumptions, biases, and perspectives, and to challenge them with evidence and logic. This self-scrutiny encourages intellectual honesty and a willingness to revise your beliefs based on new information.

Furthermore, metacognition involves seeking feedback, learning from others, and exploring different viewpoints. By being open to diverse perspectives, you can broaden your understanding and develop a more nuanced view of complex issues. This collaborative approach to learning and thinking enhances your ability to evaluate information critically and make informed decisions.

Curiosity

Curiosity serves as the driving force that motivates us to explore and understand the world around us. It propels us to delve deeper into the problems we encounter and to seek out new knowledge and insights. Curiosity encourages us to question assumptions and challenge the status quo, rather than passively accepting things as they are or dismissing them outright.

This inquisitive nature is essential for critical thinking because it stimulates our desire to learn and discover. When we are curious, we are more likely to engage in thoughtful inquiry and to connect disparate pieces of information in innovative ways. This process of connecting the dots often leads to new discoveries and creative solutions to complex problems.

Moreover, curiosity opens us up to a world of possibilities and opportunities for problem-solving and innovation. It fosters a mindset that is receptive to new ideas and perspectives, enabling us to approach challenges with a fresh and open mind. By being curious, we not only enhance our understanding of the world but also cultivate the ability to think critically and creatively.

Scepticism

Scepticism is characterised by the practice of suspending judgement. This means not rushing to conclusions but instead waiting until you have gathered a sufficient amount of relevant evidence and reasoning on an issue. To be sceptical is to withhold making decisions about the truth of claims until you have thoroughly examined the reasoning and evidence using the conceptual tools of critical thinking. It is not about automatically disbelieving everything but about maintaining a neutral stance until you can arrive at a defensible position.

This approach is more challenging than it might seem because humans are hardwired to react quickly to incoming information and situations, often making snap decisions about what to accept and how to respond. All too often, we jump to conclusions with only partial information. Our beliefs and disbeliefs are frequently knee-jerk reactions based on faulty reasoning, such as gut feelings or first impressions. Being sceptical means going through the steps of critical thinking, such as analysing meaning, logic, and argument, applying scientific methods and evidence, identifying fallacies and biases, and considering decisions and values, before forming a conclusion about a claim.

Scepticism can sometimes be perceived as a lack of commitment, and it can be difficult to maintain a neutral stance, especially when under pressure from social networks or when strong emotions are involved. This is why practising scepticism requires effort and persistence. It is important to remind yourself of the benefits of scepticism, such as avoiding gullibility and premature conclusions. Be prepared to endure the discomfort of uncertainty until you are satisfied that there is sufficient reason and evidence to take a position.

Scepticism is particularly necessary for ideas and information that are immediately appealing. Many ideas bypass our critical thinking filters because they align with our existing beliefs, such as our political, ideological, or religious views. We need to be most sceptical about ideas we want to be true. It is no accident that our interpretations are often consistent with our existing beliefs. We are almost incapable of interpreting an ambiguous event in a way that contradicts our worldview. This is because any concept, event, or interaction can be interpreted in numerous ways, allowing us to mould our interpretations to fit our belief systems. Being aware of this tendency is crucial, and it is important to critically examine and interrogate all beliefs, especially those that have not been fully scrutinised.

There are two prevalent forms of scepticism: commonsense scepticism and philosophical scepticism.

Common sense scepticism refers to the everyday practice of questioning and doubting claims until there is sufficient evidence to support them. This form of scepticism is practical and grounded in daily experiences. For example, if someone tells you that a new restaurant in town serves the best food, common sense scepticism would lead you to seek reviews, ask friends for their opinions, or try the food yourself before accepting the claim. Common sense scepticism helps us navigate the world by preventing us from being easily deceived and encouraging us to verify information before accepting it as true.

Philosophical scepticism, on the other hand, is a deeper and more systematic approach to doubt. It involves questioning the very foundations of knowledge and belief. Philosophical sceptics challenge the possibility of certainty in knowledge, asking whether we can truly know anything at all. For instance, René Descartes, a famous philosophical sceptic, doubted everything he believed until he reached the conclusion "Cogito, ergo sum" (I think, therefore I am). This form of scepticism is not about practical decision-making but about exploring the limits and nature of human knowledge. Philosophical scepticism can lead to deeper insights into the nature of reality and our understanding of it, but it can also result in radical doubt, where one questions the validity of all beliefs.

The key difference between commonsense and philosophical scepticism lies in their scope and application. Common sense scepticism is practical and applied to everyday situations, helping us make informed decisions based on evidence. For example, questioning the authenticity of a news article before sharing it is an act of common sense scepticism. Philosophical scepticism, however, is more abstract and theoretical, questioning the very possibility of knowledge itself. An example of philosophical scepticism is doubting the existence of the external world independent of our minds, as proposed by some philosophers who argue that all we know are our perceptions, which might not correspond to an external reality.

For the purposes of this book, whenever we refer to scepticism, we are referring to common sense scepticism.

Intellectual Honesty

Intellectual honesty embodies the values of objectivity, fairness, and sincerity in our thought processes.

It involves a steadfast commitment to seeking the truth, even when it is inconvenient, uncomfortable, or contrary to our pre-existing beliefs. This dedication to truth is essential for avoiding errors, biases, and fallacies in our reasoning.

Practising intellectual honesty means acknowledging our limitations and being willing to admit our mistakes. It requires us to correct our misconceptions and continuously strive for a more accurate understanding of the world. This openness to self-correction is crucial for refining our critical thinking skills and ensuring that our judgements are based on sound reasoning and reliable evidence.

Evaluating the credibility and validity of sources, arguments, and evidence is another key aspect of intellectual honesty. It involves a rigorous assessment of the information we encounter, questioning its origins, and scrutinising its reliability. By doing so, we can discern between well-founded arguments and those that are flawed or misleading.

Intellectual honesty also fosters respect and appreciation for different perspectives, opinions, and experiences, provided they are grounded in facts and logic. This respect for diverse viewpoints enhances our ability to think critically by exposing us to a broader range of ideas and encouraging us to consider alternative explanations and solutions.

Open-Mindedness

Open-mindedness embodies the willingness to consider different ideas, opinions, and perspectives, even when they challenge our own. This openness is crucial because it helps us to avoid prejudice, bias, and dogmatism in our reasoning. By embracing open-mindedness, we can acknowledge the complexity and diversity of the world we live in and appreciate the value of multiple viewpoints.

Being open-minded means actively seeking out new information and being willing to learn from others. It involves a readiness to revise our beliefs based on evidence and logic rather than clinging to preconceived notions. This flexibility in thinking allows us to adapt to new situations and to understand issues from various angles, enhancing our overall critical thinking abilities.

Open-mindedness also fosters a sense of curiosity and exploration. When we are open to different perspectives, we become more curious about the world around us and more eager to explore new ideas. This curiosity drives us to ask questions, seek out new knowledge, and engage in thoughtful discussions with others. It encourages us to connect with people who have different experiences and viewpoints, broadening our understanding and enriching our intellectual lives.

Moreover, open-mindedness promotes tolerance and empathy. By considering and respecting diverse perspectives, we develop a deeper appreciation for the experiences and opinions of others. This empathy helps us to build more inclusive and collaborative environments where different ideas can be shared and debated constructively. It also enables us to navigate conflicts and disagreements with greater understanding and respect.

Critical Thinking Skills

The second component of critical thinking is the skills component. Critical thinking is a multifaceted process that extends beyond merely possessing certain dispositions. While dispositions such as curiosity, intellectual honesty, and open-mindedness are essential, they are not sufficient on their own. To truly engage in critical thinking, one must also develop and apply a range of specific skills that enable deeper analysis and understanding. This involves the ability to interpret, analyse, evaluate, and infer, even when meanings and significance are not immediately apparent, as well as the ability to stay focused on the task at hand.

Analyse Arguments

Analysing arguments requires a comprehensive evaluation of various elements to determine the strength and validity of the argument presented. This process involves making inferences about the author's intentions and perspectives, as well as assessing the logical structure, supporting evidence, and credibility of the information sources. Additionally, it is essential to consider counterarguments to ensure a balanced and thorough analysis.

To begin with, analysing an argument necessitates a deep understanding of both deductive and inductive reasoning. Deductive reasoning involves drawing logical conclusions from general premises or principles. A premise is a statement or idea that forms the basis for an argument. It is a top-down approach where the conclusion necessarily follows from the given premises if they are true. For example, if all humans are mortal (general premise) and Socrates is a human (specific case), then Socrates is mortal (logical conclusion). This form of reasoning is crucial for testing the validity of arguments and ensuring that conclusions are logically sound.

On the other hand, inductive reasoning is the process of inferring general patterns or rules from specific observations or cases. It is a bottom-up approach where conclusions are drawn based on the evidence available, though they may not be definitively proven. For instance, observing that the sun has risen in the east every morning leads to the general conclusion that the sun always rises in the east. Another example is noticing that every time you water your plants, they appear healthier and grow faster. From these specific observations, you might infer the general rule that regular watering promotes plant health and growth. Inductive reasoning is essential for generating new knowledge, forming hypotheses, and making predictions based on empirical data, which are facts and information gathered through direct observation or experimentation.

Evaluating the quality of an argument also involves scrutinising the evidence that supports it. This means examining the relevance, reliability, and sufficiency of the evidence presented. Reliable evidence is typically derived from credible sources, such as peer-reviewed journals, reputable experts, or well-conducted studies.

It is important to assess whether the evidence is directly related to the argument and whether it is robust enough to support the claims being made.

Furthermore, analysing arguments requires considering the credibility of the information sources. This involves evaluating the expertise, objectivity, and potential biases of the sources. Credible sources are those that are recognised for their authority and reliability in the relevant field. It is also important to be aware of any potential biases that might influence the information provided, as these can affect the objectivity and accuracy of the argument.

In addition to evaluating the logic and evidence, it is essential to consider counterarguments. This involves identifying and examining alternative perspectives and objections to the argument. By addressing counterarguments, we can test the robustness of the original argument and ensure that it stands up to scrutiny. This process helps to refine our understanding and strengthen our reasoning by considering all relevant viewpoints.

As you have seen, this section describes arguments in a broad sense, as reasoned positions that require evaluating evidence, credibility, and counterarguments. This is only one aspect of understanding arguments in critical thinking. Later in this book, we will examine arguments in a more technical way, focusing on their formal structure as sets of statements where premises support a conclusion. Understanding both the broader evaluative process and the formal structure will equip you to analyse and construct arguments with greater precision.

Understand Statistics

Many of us often feel intimidated by the term 'statistics', associating it with complex formulas and calculations. However, in today's age, psychology students are not expected to memorise statistical formulas. Instead, they can rely on software such as SPSS and JAMOVI to perform these calculations automatically. This shift allows us to focus on understanding the underlying statistical concepts, which are crucial for interpreting data and making informed decisions.

Understanding statistics provides the tools needed to collect, organise, analyse, and interpret data to make informed decisions based on evidence. Statistics plays a vital role in testing hypotheses, evaluating arguments, and drawing conclusions from data, making it an indispensable skill for critical thinkers.

One fundamental concept in statistics is the normal distribution. This is a bell-shaped curve that describes how values of a variable are distributed around the mean average. A variable is any characteristic, number, or quantity that can be measured or quantified. For example, if we measure the heights of a large group of people, most heights will cluster around the average height, with fewer people being very short or very tall. The normal distribution is particularly useful for critical thinking because it allows us to estimate the probability of observing certain values, compare different groups, and identify outliers or unusual values.

By understanding the normal distribution, we can make more accurate predictions and better understand the variability within a dataset.

Another key concept is the distinction between correlation and causation. Correlation measures how two variables are related to each other, indicating whether an increase or decrease in one variable corresponds to an increase or decrease in another. For example, there might be a correlation between ice cream sales and drowning incidents, as both tend to increase during the summer. However, correlation does not imply causation, which is a stronger claim that one variable directly causes or influences another. In this case, the increase in ice cream sales does not cause more drownings; instead, both are influenced by the warmer weather. Understanding this distinction is crucial for critical thinking because it helps us explore relationships between variables, test causal hypotheses, and avoid logical fallacies or false assumptions. Just because two variables are correlated does not mean that one causes the other; there could be other underlying factors at play.

In addition to these concepts, critical thinkers must be adept at evaluating the quality of statistical evidence. This involves assessing the methods used to collect and analyse data, the sample size, and the potential biases that might affect the results. For example, if a study claims that a new drug is highly effective based on a small sample size of only 10 participants, a critical thinker would question the reliability of these results. They would consider whether the sample size is large enough to draw meaningful conclusions and whether the study design might have introduced any biases. By critically evaluating statistical evidence, we can determine the reliability and validity of the conclusions drawn from the data.

Furthermore, understanding statistics enables us to communicate findings effectively. Being able to interpret and present data clearly and accurately is essential for making persuasive arguments and informed decisions. This skill is particularly important in psychology, where data-driven insights are crucial for understanding human behaviour and developing effective interventions.

Evaluate Experimental Designs

Experimental designs are structured methods for conducting controlled experiments to test the effects of one or more factors on an outcome. Understanding these designs is essential because they help us establish causality, control for confounding variables, and measure the significance of results. A confounding variable is an external factor that can influence the experimental variables, potentially leading to incorrect conclusions about the relationship between them.

One common type of experimental design is the randomised controlled trial (RCT). In an RCT, participants are randomly assigned to either the treatment group or the control group. This randomisation helps ensure that any differences observed between the groups are due to the treatment itself and not other factors. For example, in a clinical trial testing a new medication, randomising participants helps control for potential confounding variables such as age, gender, and health status, which might otherwise influence the results. RCTs are considered the gold standard for testing the efficacy of interventions because they provide strong evidence of causality.

Another important experimental design is the factorial design. This approach allows researchers to test the effects of multiple factors simultaneously and to explore interactions between them. For instance, a study might investigate the combined effects of diet (factor 1) and exercise (factor 2) on weight loss (outcome variable) by assigning participants to different groups based on their diet and exercise regimens. Factorial designs are valuable because they can reveal how different factors (e.g., diet and exercise) interact and influence outcomes (e.g., weight loss), providing a more comprehensive understanding of complex phenomena.

Quasi-experiments are also widely used, particularly when randomisation is not feasible or ethical. In quasi-experiments, participants are not randomly assigned to groups, but the researchers still attempt to control for confounding variables through other means. For example, a study examining the impact of a new educational program might compare outcomes between schools that adopt the program and those that do not while accounting for differences in student demographics and school resources. Although quasi-experiments are less rigorous than RCTs, they can still provide valuable insights, especially in real-world settings where randomisation is impractical.

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

1.4. THE OBSERVE FRAMEWORK

By Marc Chao

In today's information age, individuals are constantly exposed to a vast amount of unverified information from various sources, including social media, news outlets, and even personal conversations. Some of this information can be sensational, emotional, or even outrageous, prompting quick reactions, hasty generalisations, and conclusions. However, such rapid judgements or ill-informed hypotheses can often lead to misunderstanding or misinformation. To navigate this flood of information and make informed decisions, it is essential to apply a critical thinking approach.

The OBSERVE framework (Figure 1.4.1) is a structured critical thinking method designed to help individuals systematically assess and evaluate hypotheses in response to observed phenomena or received information. It guides you through a process of inquiry and analysis, encouraging a thorough examination of the evidence, reflection on personal biases, and careful consideration of various hypotheses to determine which one is better supported by the available information.

The framework comprises the following seven steps:

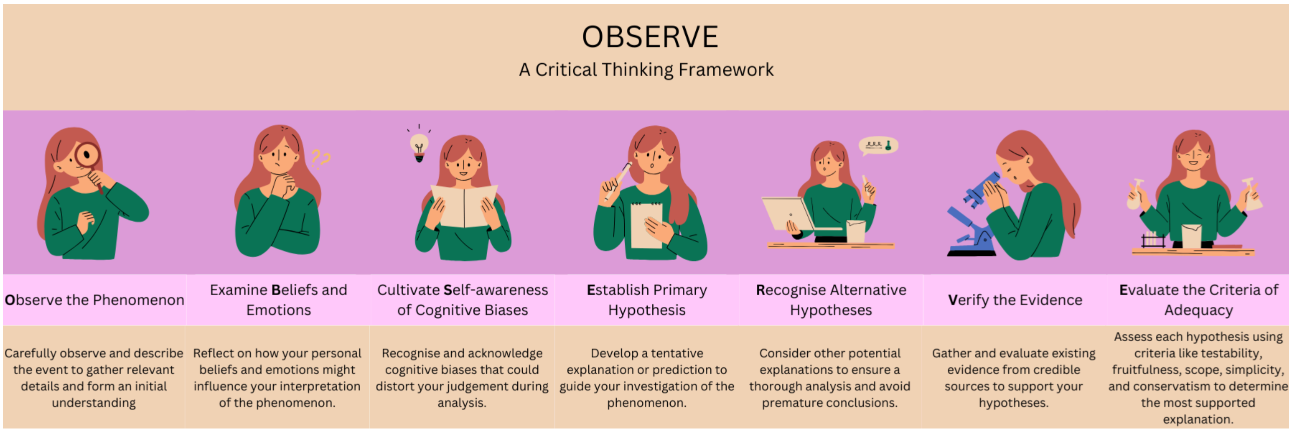


Figure 1.4.1. The OBSERVE framework for critical thinking by Marc Chao is used under a

1. **O**bserve the Phenomenon
2. Examine **B**eliefs and Emotions
3. Cultivate **S**elf-awareness of Cognitive Biases
4. **E**stablish Primary Hypothesis
5. **R**ecognise Alternative Hypotheses

6. **V**erify the Evidence
7. **E**valuate the Criteria of Adequacy.

1. Observe the Phenomenon

The first step in the OBSERVE framework is to observe the phenomenon. This involves carefully noticing and describing the event or situation you are investigating. For instance, you may observe that your performance on a recent test was subpar, potentially due to feeling unusually fatigued before the exam. This differed from a previous test in which you performed well after having had a restful night's sleep. This observation may lead you to hypothesise that sleep deprivation could influence test performance.

In the realm of psychology and other sciences, observation is the foundation upon which all subsequent steps are built. Before formulating hypotheses or designing experiments, scientists must first observe a phenomenon that piques their curiosity or raises questions. This initial observation phase is essential because it allows researchers to identify patterns, anomalies, or interesting behaviours that warrant further investigation. For instance, a psychologist might notice that individuals who report high levels of stress also seem to have difficulty concentrating. This observation could lead to a more structured investigation into the relationship between stress and cognitive function.

Observation is not a passive activity; it requires active engagement and a keen eye for detail. Scientists must be meticulous in their observations, taking note of all relevant factors and conditions. This might involve recording behaviours, measuring physiological responses, or documenting environmental conditions. Inaccurate or incomplete observations can lead to flawed hypotheses and unreliable results.

Furthermore, observation is an ongoing process throughout the research. As experiments are conducted and data is collected, researchers continue to observe and refine their understanding of the phenomenon. They might notice new patterns or unexpected results that prompt further investigation. This iterative process of observation and hypothesis testing is at the heart of the scientific method, driving the advancement of knowledge.

2. Examine Beliefs and Emotions

The next step is to examine your own beliefs and emotions. This involves being aware of how your personal beliefs, emotions, and biases might shape your observations of the phenomenon. For example, if you believe that sleep deprivation significantly impacts cognitive performance, you may be more inclined to focus on aspects of the phenomenon that support this belief, such as feeling tired or struggling to concentrate. On the other hand, you may overlook or downplay evidence, such as social or environmental factors, that may contradict this view. By recognising the potential influence of your beliefs and emotions, you can approach your observations more objectively, which helps in forming a more balanced hypothesis.

Emotions and beliefs play a crucial role in how we observe and interpret phenomena. Strong emotions, such as excitement, frustration, or anxiety, can distort our observations and lead us to focus on certain details while neglecting others. For instance, if you have a personal interest in a new intervention or theory, you might be more likely to notice positive signs and ignore any negative ones. Alternatively, if you are sceptical about a certain concept, your emotions may cause you to focus on its flaws while dismissing any potential benefits. Recognising the influence of emotions and beliefs is key to maintaining objectivity and ensuring that your observations are not unduly affected by personal biases.

In addition to emotions, our existing beliefs can lead us to make biased observations. This tendency, known as motivated reasoning, involves interpreting new information in a way that aligns with our pre-existing views. For instance, if you strongly believe in the benefits of a specific diet, you might selectively focus on any positive outcomes related to it, while ignoring other factors that could explain those results. This tendency to interpret observations through the lens of our beliefs can limit our ability to consider alternative explanations and hinder the development of objective hypotheses.

To mitigate these biases, it is important to practice self-awareness and critical reflection throughout the observation process. Begin by acknowledging your beliefs and emotions and how they may be influencing what you observe. Ask yourself questions like, "Am I focusing on evidence that confirms my beliefs?" or "How might my emotional state be shaping my observations?". By actively questioning your own thought processes, you can identify and counteract potential biases, helping to ensure that your observations are more accurate and reflective of the phenomenon itself.

3. Cultivate Self-Awareness of Cognitive Biases

Building on the previous step, this involves cultivating a deeper self-awareness of your cognitive biases. Cognitive biases are systematic patterns of deviation from rationality or logical judgement.

After observing a phenomenon, it is natural to form initial impressions and hypotheses. However, these initial thoughts are often influenced by cognitive biases. For example, if you have a tendency to believe that a particular teaching method is ineffective, you might be more likely to notice and emphasise instances where students struggle under this method, while overlooking cases where they succeed. This selective attention can lead to confirmation bias, where we favour information that supports our existing views and ignore or downplay information that contradicts them. Similarly, the anchoring bias can also play a role. If you initially observe poor performance with a specific method, you might anchor your evaluation to that first impression, making it harder to adjust your opinion even as new, more favourable evidence emerges. By recognising these biases, you can take steps to reduce their impact, such as seeking out disconfirming evidence or using blind analysis techniques.

One effective strategy is to actively seek out disconfirming evidence. Rather than focusing solely on data that supports your hypothesis, look for information that challenges it. This approach helps to balance your perspective and reduce the risk of confirmation bias. For instance, if you believe that a new educational program is effective, examine cases where it did not produce the expected results and consider alternative explanations.

Another strategy is to involve others in your research process. Collaborating with colleagues or seeking feedback from peers can provide fresh perspectives and help identify biases you might have overlooked. Others can offer critical insights and challenge your assumptions, leading to a more robust and objective analysis.

4. Establish Primary Hypothesis

Once a phenomenon has been thoroughly observed and you have reflected on your own beliefs, emotions, and biases, the next step is to establish a primary hypothesis. This is a tentative explanation or prediction that you will test through further investigation. This is where the interplay between observation and creativity becomes evident. Observing a phenomenon sparks curiosity and prompts questions, leading researchers to think creatively about potential explanations. For example, after observing that sleep-deprived individuals perform poorly on cognitive tasks, a psychologist might hypothesise that sleep deprivation impairs memory consolidation processes in the brain.

A well-crafted hypothesis should be specific and measurable. For example, instead of a vague hypothesis like "sleep affects cognitive performance", a more precise hypothesis would be "sleep deprivation of less than five hours per night for a week negatively impacts short-term memory performance in adults". This specificity allows for clear operational definitions and measurable outcomes, making it easier to design experiments and analyse results.

The process of establishing a primary hypothesis often involves reviewing existing literature and theories related to your observed phenomenon. This background research helps ensure that your hypothesis is grounded in current knowledge and addresses gaps or unresolved questions in the field. For instance, if previous studies have shown mixed results on the impact of sleep deprivation on cognitive performance, your hypothesis might aim to clarify these inconsistencies by focusing on a specific aspect, such as short-term memory.

In addition to being specific and measurable, a good hypothesis should be falsifiable. This means that it should be possible to prove the hypothesis wrong through empirical testing. Falsifiability is a cornerstone of the scientific method because it allows for hypotheses to be rigorously tested and potentially refuted, leading to a more robust understanding of the phenomenon.

Once you have formulated your primary hypothesis, it serves as a guiding framework for your research. It informs the design of your experiments, the selection of variables, and the methods of data collection. For example, if your hypothesis is that sleep deprivation negatively affects short-term memory, you might design an experiment where participants are assigned to different sleep conditions and their memory performance is assessed using standardised tests.

Establishing a primary hypothesis also helps in setting clear objectives and expectations for your research. It provides a benchmark against which you can compare your results and determine whether your hypothesis is supported or refuted. This clarity is essential for maintaining focus and direction throughout the research process.

5. Recognise Alternative Hypotheses

Critical thinking involves considering multiple perspectives, so it is important to recognise alternative hypotheses. These are other possible explanations for the phenomenon you are studying. Considering alternative hypotheses helps ensure that you do not prematurely settle on a single explanation and encourages a more comprehensive analysis of the data.

Recognising alternative hypotheses requires a creative and open-minded approach and a willingness to entertain different possibilities. For example, if your primary hypothesis is that sleep deprivation negatively affects cognitive performance, an alternative hypothesis might be that the observed cognitive decline is due to stress rather than sleep deprivation. This alternative explanation prompts you to consider additional variables and control for potential confounding factors in your study.

Exploring alternative hypotheses can also lead to a more nuanced understanding of the phenomenon. For instance, you might hypothesise that both sleep deprivation (factor 1) and stress (factor 2) contribute to cognitive decline, but in different ways or to varying degrees. This integrated perspective can help you design more comprehensive experiments that account for multiple factors and their interactions.

To effectively recognise alternative hypotheses, it is helpful to engage in brainstorming sessions and discussions with colleagues or peers. Collaborative thinking can generate a wider and more creative range of ideas and perspectives, helping to identify potential explanations that you might not have considered on your own. Additionally, reviewing diverse literature and case studies can provide insights into different factors that could influence the phenomenon.

Recognising alternative hypotheses also enhances the credibility and rigour of your research. It demonstrates a thorough and critical approach to inquiry, showing that you have considered multiple angles and potential explanations. This comprehensive analysis is essential for building a robust body of evidence and advancing knowledge in your field.

6. Verify the Evidence

The next step is to verify the evidence by gathering preliminary information to assess whether your hypotheses are worth testing. At this stage, instead of collecting primary data through experiments or surveys, you search for existing evidence that supports your primary and alternative hypotheses (we will see how these two evidence types contrast and challenge each other in Step 7). This involves reviewing relevant literature, conducting internet searches, and examining credible sources to evaluate whether there is sufficient empirical support for the hypotheses you intend to test.

Verifying evidence at this stage is crucial for ensuring that your hypotheses are grounded in reliable information and that your planned study will contribute meaningfully to the field. By gathering information from reputable sources, such as peer-reviewed journal articles or trusted academic publications, you can gauge whether your hypotheses are based on sound findings. On the other hand, sources such as social media posts or non-expert blogs may not provide reliable evidence, and it is important to critically evaluate these materials before considering them in your planning process.

During the verification process, it is important to assess the quality of the evidence you gather. You should scrutinise the credibility, methodology, and potential biases of the sources you encounter. For example, articles from peer-reviewed journals or research conducted by established experts in the field are generally considered more reliable than studies from questionable outlets. By verifying the credibility of your sources, you can identify any potential flaws or biases in the existing literature that might affect the hypotheses you are considering.

In addition to evaluating the credibility of your sources, it is important to assess the relevance and applicability of the evidence. Are the studies or articles you review directly related to your research question? Do they use appropriate methodologies and cover similar contexts or populations? For example, if you are considering the hypothesis that sleep deprivation negatively affects cognitive performance, you would want to review studies that examine similar variables, such as cognitive tests conducted on sleep-deprived individuals and assess how closely these studies align with your intended research design.

You should also examine the strength of the evidence you find. This includes evaluating the sample size, research design, and any statistical analysis used in prior studies. Strong evidence typically comes from well-designed studies with large, representative samples, appropriate controls, and robust statistical methods. For example, if you are testing the impact of sleep deprivation on cognitive performance, studies with carefully controlled sleep conditions and reliable cognitive tests would provide a more solid foundation for your hypothesis than studies with small sample sizes or poorly controlled variables.

Another important aspect of verifying evidence is to assess whether the findings from existing research can be replicated. If similar studies have consistently supported your hypothesis, this strengthens the case for further investigation. Conversely, if the evidence is inconsistent or there are significant challenges to replicating the findings, you might reconsider or refine your hypotheses. Replicability is a hallmark of reliable evidence and gives you greater confidence in the foundation upon which you are building your research.

Finally, it is essential to consider the ethical dimensions of the evidence you gather. While you are not yet conducting experiments or collecting primary data, you should still ensure that any existing research adheres to ethical standards, such as informed consent, privacy protection, and the minimisation of harm. Reviewing ethical considerations in prior studies can provide insight into how similar research has been conducted responsibly and can help inform your research design when the time comes to gather your data.

7. Evaluate the Criteria of Adequacy

Evaluating the criteria of adequacy for each hypothesis is a crucial step in the critical thinking process. This evaluation ensures that the hypotheses you consider are not only plausible but also robust and useful for advancing knowledge. The criteria of adequacy provide a systematic way to assess the strengths and weaknesses of each hypothesis, guiding you toward the most reliable and informative explanations. By rigorously applying these criteria, you can enhance the quality of your research and avoid common pitfalls such as confirmation bias or overcomplication.

In this step, you will use the evidence gathered in Step 6 of the OBSERVE framework to compare the hypotheses. The preliminary evidence from literature reviews, internet searches, and other reliable sources provides the foundation upon which you can critically compare the hypotheses.

The criteria of adequacy include testability, fruitfulness, scope, simplicity, and conservatism. Each of these criteria serves a specific purpose in evaluating hypotheses, helping to ensure that they are scientifically sound and practically useful.

Criterion 1: Testability

Testability is a fundamental principle in scientific inquiry. For a hypothesis to be useful, it must be testable, meaning it can be investigated through observation or experimentation and potentially proven wrong. This ensures that hypotheses are based on observable, measurable phenomena, not on vague or untestable ideas. For example, the hypothesis “sleep deprivation negatively affects cognitive performance” is testable because it can be studied through controlled experiments and cognitive tests.

A testable hypothesis allows researchers to design experiments to confirm or challenge it. This involves identifying specific variables, defining them clearly, and developing ways to measure their impact. In the sleep deprivation example, researchers might manipulate how much sleep participants get and then measure their cognitive performance using standard tests. If participants’ performance consistently declines with less sleep, the hypothesis is supported; if there is no change, the hypothesis may be disproven.

Testability also includes the idea of falsifiability. A hypothesis is falsifiable if there are potential observations or experiments that could prove it wrong. For example, the hypothesis “all swans are white” is falsifiable because it can be tested by looking for swans. If even one black swan is found, the hypothesis is disproven. Falsifiability is crucial for scientific progress because it allows theories to be refined over time. On the other hand, a non-falsifiable hypothesis cannot be tested or proven wrong. For instance, the hypothesis that "an invisible, undetectable force influences our thoughts" is non-falsifiable because there is no way to test or disprove it. Hypotheses that cannot be tested or falsified are not helpful in scientific research, as they do not contribute to advancing knowledge.

Criterion 2: Fruitfulness

Fruitfulness refers to a hypothesis's ability to lead to new insights, predictions, and research questions. A fruitful hypothesis opens up new areas of exploration and helps expand scientific knowledge. For example, the hypothesis that sleep deprivation affects cognitive performance could lead to studies on which specific cognitive functions are impacted, the underlying brain processes, and ways to reduce the effects.

A fruitful hypothesis does not just explain what we observe; it also predicts new phenomena that can be tested. These predictions guide future research and help build a fuller understanding of the topic. For instance, if sleep deprivation is found to affect memory, researchers might predict that it also impacts attention and decision-making, leading to new studies in those areas.

Fruitfulness is also linked to heuristic value, which is a hypothesis's ability to inspire further research and discovery. A hypothesis with high heuristic value encourages scientists to explore new ideas, develop innovative methods, and expand knowledge. This ongoing cycle of generating and testing hypotheses is essential for scientific progress.

Criterion 3: Scope

Scope refers to the range and general applicability of a hypothesis. A hypothesis with a broad scope applies to many situations and phenomena, making it more useful for understanding complex issues. For example, a hypothesis that sleep deprivation affects not just cognitive performance, but also emotional regulation and physical health has a broader scope than one that focuses only on cognitive performance.

A broad hypothesis can explain a wide range of observations, offering a more complete understanding of the phenomenon. This is especially important in fields like psychology, where many factors interact to affect behaviour and mental processes. For instance, a broad hypothesis about sleep deprivation could include its effects on mood, stress, immune function, and overall well-being.

However, it is important to strike a balance between scope and specificity. While a broad hypothesis may provide more insight, it must still be clear and testable. Researchers should aim to develop hypotheses that are both wide-ranging and grounded in evidence, ensuring they can be tested rigorously and lead to meaningful conclusions.

Criterion 4: Simplicity

Simplicity, or parsimony, refers to the preference for hypotheses that are clear and straightforward, without unnecessary complexity. This idea follows Occam's Razor, which suggests that when faced with competing hypotheses, the one with the fewest assumptions is usually the best choice. A simple hypothesis is easier to test, understand, and communicate, making it more practical for scientific research.

For example, the hypothesis "sleep deprivation negatively affects cognitive performance" is simpler than a more complicated hypothesis that involves many interacting factors without clear definitions. Simplicity does not mean oversimplifying; it means focusing on the most important relationships and mechanisms while removing unnecessary details.

A simple hypothesis is easier to test and falsify because it has fewer variables and assumptions. This clarity makes it easier to design experiments and interpret results. However, simplicity should not sacrifice the ability to explain the phenomenon. The hypothesis must still accurately account for the observed data and offer meaningful insights.

Criterion 5: Conservatism

Conservatism refers to how well a hypothesis aligns with existing knowledge and theories. A conservative hypothesis fits with what is already known and does not require drastic changes to well-established scientific ideas. This ensures that new hypotheses build on existing knowledge, rather than contradicting or ignoring it.

For example, a hypothesis that sleep deprivation affects cognitive performance is conservative if it aligns with existing research on how sleep impacts brain function. On the other hand, a hypothesis proposing a completely new, unsupported mechanism would be less conservative and would require more rigorous testing.

Conservatism does not mean rejecting new ideas; rather, it highlights the importance of building on what we already know. New hypotheses should refine and extend current theories, integrating new findings into the larger body of scientific knowledge. This approach helps maintain the stability and reliability of science while still allowing for new discoveries.

Applying the OBSERVE Framework

The OBSERVE framework is a powerful technique that fosters critical thinking by providing a structured approach to analysing and understanding phenomena. It can be applied whenever someone encounters a new or complex situation, guiding them through systematic observation, reflection on personal biases, hypothesis formulation, consideration of alternative explanations, evidence verification, and rigorous evaluation of hypotheses. By employing the following steps, individuals can develop well-founded conclusions and make informed decisions, enhancing their ability to navigate and interpret the world around them.

Step 1: Observe the Phenomenon: Begin by carefully observing and describing the event or situation you are investigating. This step involves gathering initial data and noting relevant details to form a clear understanding of the phenomenon.

Step 2: Examine Beliefs and Emotions: Reflect on your own beliefs and emotions that might influence your interpretation of the phenomenon. Recognising how your personal perspectives can affect you helps maintain objectivity.

Step 3: Cultivate Self-awareness of Cognitive Biases: Develop a deeper awareness of your cognitive biases. These biases can skew your judgement, so acknowledging them is crucial for unbiased analysis.

Step 4: Establish Primary Hypothesis: Formulate a primary hypothesis, which is a tentative explanation or prediction about the phenomenon. This hypothesis will guide your investigation and focus your efforts on testing its validity.

Step 5: Recognise Alternative Hypotheses: Consider other possible explanations for the phenomenon. Recognising alternative hypotheses ensures a comprehensive analysis and prevents premature conclusions.

Step 6: Verify the Evidence: Search for existing evidence that supports your primary and alternative hypotheses. This step involves gathering and assessing the quality of evidence through reviewing relevant literature, conducting internet searches, and examining various sources of information.

Step 7: Evaluate the Criteria of Adequacy: Compare and contrast the primary hypothesis against the alternative hypotheses using the Criteria of Adequacy (Table 1.4.1), as follows:

Table 1.4.1. Applying the Criteria of Adequacy

	Using the evidence gathered in Step 6, which hypotheses (primary or alternative):
Testability	Is more amenable to empirical testing and falsification?
Fruitfulness	Is more likely to generate new insights and predictions?
Scope	Is more broadly applicable to a range of situations?
Simplicity	Is more straightforward, with fewer unnecessary complexities?
Conservatism	Aligns more closely with established knowledge and theories?

The hypothesis that satisfies more of these criteria is considered better supported by the evidence. Therefore, when encountering potentially alarming or unverified information, whether online or from a personal source, it is advantageous to apply the OBSERVE framework. This approach encourages critical thinking and helps assess the credibility and accuracy of the information.

1.5. SUMMARY

By Marc Chao

Summary

Critical thinking is the disciplined process of analysing, synthesising, and evaluating information to make informed decisions, solve problems, and understand the world. It involves questioning assumptions, evaluating evidence, and considering alternative perspectives, which helps individuals navigate biases like confirmation bias and misinformation. By actively addressing biases and fostering openness to new information, critical thinking enables proactive decision-making and goes beyond common sense. In today's fast-paced, media-driven world, critical thinking is vital for discerning truth amidst overwhelming information and misinformation. Its significance is evident in addressing societal challenges like vaccine hesitancy and the spread of false narratives. As an essential 21st-century skill, critical thinking combines problem-solving, creativity, and decision-making to empower individuals to thrive in complex environments.

Creativity and critical thinking are deeply interdependent, with creativity focused on generating novel ideas and perspectives, while critical thinking involves analysing, evaluating, and applying them. Together, they are indispensable for effective problem-solving, communication, and innovation, as seen in fields like songwriting, investigative journalism, and experimental research. Creativity fosters imaginative approaches to challenges, and critical thinking ensures these ideas are rigorously assessed and practically applied. Overcoming the misconception that these skills are incompatible requires understanding their complementary nature, relying on both hemispheres of the brain working together. To enhance creativity, individuals should seek diverse perspectives, challenge assumptions, and practice generating ideas without judgement, as exposure to new environments and ideas fuels creativity.

Becoming a critical thinker requires a purposeful and disciplined approach to evaluating beliefs, claims, or decisions against evidence and context. It calls for specific dispositions, such as curiosity, open-mindedness, metacognition, intellectual honesty, and scepticism, alongside skills like analysing arguments, understanding statistics, and evaluating experimental designs. These

dispositions encourage reflective and open engagement with ideas, while the skills enable the assessment of evidence, identification of logical connections, and formulation of informed judgements. For example, analysing arguments involves examining evidence, logic, and counterarguments, while understanding statistics clarifies concepts like correlation versus causation.

The OBSERVE framework offers a structured approach to critical thinking, designed to systematically evaluate and verify hypotheses in an era of abundant and often misleading information. This method consists of seven steps: observing phenomena to gather data, examining beliefs and emotions to reduce bias, cultivating awareness of cognitive biases, formulating a primary hypothesis, considering alternative hypotheses, verifying evidence through credible sources, and evaluating hypotheses using the Criteria of Adequacy: testability, fruitfulness, scope, simplicity, and conservatism. By addressing biases, promoting comprehensive analysis, and emphasising evidence-based reasoning, this framework fosters objectivity and equips individuals to navigate complex challenges effectively.

CHAPTER 2: ARGUMENTS AND CRITICAL THINKING

In the realm of psychology and across numerous disciplines, the ability to form, analyse, and evaluate arguments is an indispensable skill. While the term "argument" might conjure images of heated debates or quarrels, in the context of critical thinking and academic study, an argument refers to a structured presentation of reasoning and evidence to support a conclusion. Developing proficiency in argumentation is not merely an exercise in rhetoric; it is a fundamental tool for understanding complex issues, evaluating evidence, and making informed decisions.

Arguments form the backbone of scientific inquiry, policy debates, and everyday decision-making. In psychology, for example, researchers rely on arguments to interpret data, justify methodologies, and draw meaningful conclusions from experimental results. Whether discussing cognitive biases, analysing behavioural trends, or evaluating therapeutic interventions, psychologists must build arguments that are clear, logical, and well-supported by evidence.

This chapter begins by demystifying the concept of an argument, distinguishing it from everyday disagreements or mere assertions. Through practical examples and structured explanations, readers will learn to identify the core components of an argument: premises (the reasons provided) and conclusions (the claims supported by those reasons). Additionally, the chapter explores how arguments can be strengthened or undermined, introducing key principles such as validity, soundness, and cogency.

Beyond constructing arguments, this chapter also emphasises the importance of recognising implicit premises, differentiating between inductive and deductive reasoning, and understanding common fallacies. These skills not only enhance academic performance but also equip readers to navigate the information-saturated modern world, where misinformation, biases, and logical errors are abundant.

At its core, this chapter serves as both a guide and a toolkit for sharpening critical thinking skills. It encourages readers to approach claims with curiosity and scepticism, to ask probing questions, and to construct arguments that are both persuasive and intellectually rigorous. Whether you are analysing a psychological study, evaluating a social policy, or engaging in everyday reasoning, the ability to think critically and argue effectively is an invaluable asset.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand the structure of arguments:** Define key components of an argument, including premises, propositions, and conclusions, and explain their roles in logical reasoning.
- **Differentiate between deductive and inductive arguments:** Identify the differences between deductive and inductive reasoning and understand when each is appropriate to use.
- **Evaluate argument validity and soundness:** Assess whether an argument's premises logically support its conclusion and determine if the premises are true.
- **Identify implicit premises:** Recognise and articulate unstated assumptions that are necessary for an argument to be valid.
- **Recognise common argument patterns:** Understand standard argument forms such as Modus Ponens, Modus Tollens, and disjunctive syllogisms, and identify common structural fallacies.
- **Distinguish between rational and empirical propositions:** Explain the difference between premises based on logical reasoning and those grounded in observable evidence.
- **Analyse normative and descriptive conclusions:** Differentiate between claims about how things are (descriptive) and how they should be (normative).
- **Identify and avoid common fallacies:** Recognise common logical fallacies in arguments and understand why they undermine reasoning.
- **Construct clear and persuasive arguments:** Build well-structured arguments that are logically valid and supported by sound premises.
- **Critically assess arguments:** Apply critical thinking skills to analyse and evaluate the quality and strength of arguments in both academic and real-world contexts.

2.1. PROPOSITIONS AND ARGUMENTS

By Stephanie Gibbons and Justine Kingsbury, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Learning to form arguments is a fundamental aspect of developing critical thinking skills, which are essential in psychology and many other fields. While it might seem like an exercise more suited to an English class, the ability to construct and analyse arguments is crucial for evaluating evidence, making informed decisions, and understanding complex issues. In psychology, forming arguments helps us to systematically assess theories, interpret research findings, and engage in scientific debates. This process involves not just presenting information but also providing logical reasoning and evidence to support our conclusions.

This is not an "English" class because our focus is not on language arts or literary analysis. Instead, we are concerned with the principles of logical reasoning and the application of these principles to psychological concepts. By learning to form well-structured arguments, we enhance our ability to think critically about psychological phenomena, identify cognitive biases, and improve our research methodologies. This skill set is invaluable for conducting rigorous scientific research and for applying psychological knowledge in practical, real-world situations.

In this book, we use the term "argument" in a more technical sense. In everyday language, any kind of verbal disagreement is often called an argument. For example, consider a discussion between two colleagues about the effectiveness of a new workplace policy. One colleague insists that the policy has significantly improved productivity, while the other disagrees. Their conversation might go like this:

Colleague A:	The new policy has really boosted our productivity.
Colleague B:	I don't think it has.
Colleague A:	Yes, it has.
Colleague B:	No, it hasn't.
Colleague A:	Yes, it has.

This exchange, although a verbal disagreement, does not constitute an argument in the sense we will be using the term. They were making contradictory statements without providing any supporting reasons.

An argument, as we will define it, is a set of statements that includes a conclusion and reasons to support that conclusion. If either colleague had provided reasons for their claims, they would have had the beginnings of an argument. For instance, one colleague might believe the policy is effective because they have observed a decrease in project completion times. The other might disagree because they have noticed an increase in employee stress levels.

Colleague A:	The new policy has reduced project completion times by 20%, indicating improved productivity.
Colleague B:	The new policy has increased employee stress levels, which could negatively impact long-term productivity.

Notice that they are no longer just contradicting each other ("Yes, it has!" "No, it hasn't!"). Instead, they are providing reasons to support their assertions, thus forming arguments.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=829#h5p-2

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=829#h5p-3

Not all statements or writings constitute an argument. A piece of speech or writing qualifies as an argument only if the speaker or writer aims to persuade someone of a particular point by offering reasons to support its validity.

Propositions and Arguments

Premise

The term "premise" refers to a reason or basis for believing or concluding something. Premises are statements that form the foundation of an argument. For example, if I observe that I am getting wet while standing outside, this can serve as a premise to support the conclusion that it is raining. In this case, both statements, "I am getting wet" and "It is raining", are propositions about the state of the world, but they serve different roles. The first statement, "I am getting wet", provides a reason to believe the second statement, "It is raining". Therefore, the first proposition is a premise, and the second is a conclusion.

It is important to note that not all propositions are premises. For a proposition to be a premise, it must support another proposition, which is the conclusion. Only when a proposition is used in this supportive role does it become a premise.
Suppose someone says, "The roads are slippery because it rained last night".

- In this statement, the proposition "It rained last night" functions as a premise, because it provides the reason for believing the conclusion that "The roads are slippery".
- On the other hand, if the proposition "It rained last night" is simply mentioned in isolation as a piece of information, such as in the statement "It rained last night, and I enjoyed listening to the sound of the rain", it is not a premise because it is not being used to support any conclusion.

Thus, whether a proposition is a premise depends on its role within the context of an argument. Only when it supports a conclusion does it become a premise.

Proposition

A "proposition" or "declarative statement" is a sentence that makes a claim about something that can be either true or false. The terms "declarative statement", "proposition", and "statement" are often used interchangeably.

Critical thinking involves figuring out what to believe, what we know, what is true and reasonable, and what is worthwhile. Propositions are the statements that express or summarise our beliefs and knowledge, unlike other statements that ask questions or give directions. This book will focus on propositions and not cover other types of statements. Make sure you understand what a proposition is and why it is important to critical thinking before moving on.

Language is the tool we use to accomplish almost everything in our lives. The parts of our language that express our thoughts, beliefs, and knowledge are called propositions. To become skilled critical thinkers, it is essential to understand these types of statements, their functions, how to evaluate them critically, and how they are used in reasoning (i.e., arguments).

The key feature of a proposition, unlike other types of statements, is that it can be "true" or "false". Questions and commands are not the types of statements that can be true or false; they serve different purposes.

As discussed in the "Premise" section above, propositions can serve various roles, such as premises or conclusions in arguments. For example, the statement "I am getting wet" is a proposition because it makes a claim that can be true or false. If this claim is true, the proposition can serve as a premise to build an argument that supports another proposition, called a "conclusion", such as "It is raining". In this way, conclusions, like premises, are also propositions whose truth depends on the truth of the premises provided. Organising sentences with premises and a conclusion is known as an "argument".

Argument

An argument is a group of propositions or statements that form a piece of reasoning. Reasoning involves using certain information (called premises) to derive other information (called conclusions) through inference. An inference is the mental process of connecting premises to reach a conclusion that logically follows from them.

In an argument, one or more propositions serve as premises or reasons, and one proposition serves as the conclusion. An argument must have at least two propositions or statements, but it often includes more. For example:

Premise 1:	All humans are mortal.
Premise 2:	Socrates is a human.
Conclusion:	Therefore, Socrates is mortal.

Here, the premises are logically connected and lead to the conclusion.

A single statement cannot be a complete argument. For example:

Statement:	Socrates is mortal.

This is not an argument because it presents only a conclusion without any premises to support it.

Statements within an argument must be related in a way that establishes a reason-and-conclusion relationship. If the propositions do not relate to each other in this way, they do not form an argument. For example:

Statement 1:	It is sunny today.
Statement 2:	My favourite colour is purple.

These statements do not relate to each other in a way that establishes a reason-and-conclusion relationship, so they do not form an argument.

The purpose of an argument is to convince others of the conclusion. Just as one should not hold a belief without a reason, one cannot have a conclusion without premises. The sentences in an argument are all propositions or declarative statements, each serving different roles within the argument structure. The role of the premises is to support or justify the conclusion. While an argument can have only one conclusion, it can have multiple premises. In an argument, premises are assumed to be true, though this does not guarantee their truth; they are not justified within the argument itself. Other arguments likely justify these premises, where they would serve as conclusions.

The effectiveness of premises in supporting the conclusion and the credibility of their connection to the conclusion are critical aspects of analysing arguments.

Chapter Attribution

Content adapted, with editorial changes, from:

How to think critically (2024) by Stephanie Gibbons and Justine Kingsbury, University of Waikato, is used under a CC BY-NC licence.

2.2. CONSTRUCTING AN ARGUMENT

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

All statements in an argument are propositions, which means they are statements that can be either true or false. To make a proposition credible or believable, it needs the support of a strong argument. While premises within an argument serve as reasons to support the conclusion, they are not the primary focus of the argument itself. The main goal of an argument is to prove its conclusion, assuming that the premises are already justified or self-evident. If the premises need further support, they must act as conclusions in separate arguments with their own supporting premises, creating a continuous cycle of reasoning. In most arguments, however, premises are presented as either self-evident or already established. If this is not the case, the person making the argument can either provide additional support for the premises or leave it to the audience to assess their validity.

This focus on propositions is crucial because our thinking primarily occurs through words, structured into sentences or statements. While emotions and mental imagery play a role, our reasoning and communication depend on language, particularly on declarative sentences or propositions. Propositions are essential building blocks of reasoning, as they are the means through which we express and evaluate truth. In arguments, propositions are grouped together: some serve as supporting reasons (premises), while others are the main focus (conclusions). This relationship between premises and conclusions forms the foundation of critical thinking and logical reasoning.

To summarise (Figure 2.2.1), propositions are a specific type of statement that makes an assertion about something being true or false. This distinguishes them from other types of statements, such as interrogative statements (questions) or imperative statements (commands or requests). Propositions are grouped in arguments to justify a conclusion, with premises acting as the supporting reasons. Premises can take different forms: they may describe the nature of things (rational premises), provide evidence from observations (empirical premises, including scientific evidence), or rely on definitions (definitional premises). Occasionally, rhetorical questions or commands might be used in arguments, though they are less formal and less convincing. Ultimately, propositions play a central role in reasoning, as they allow us to construct, evaluate, and communicate ideas effectively.

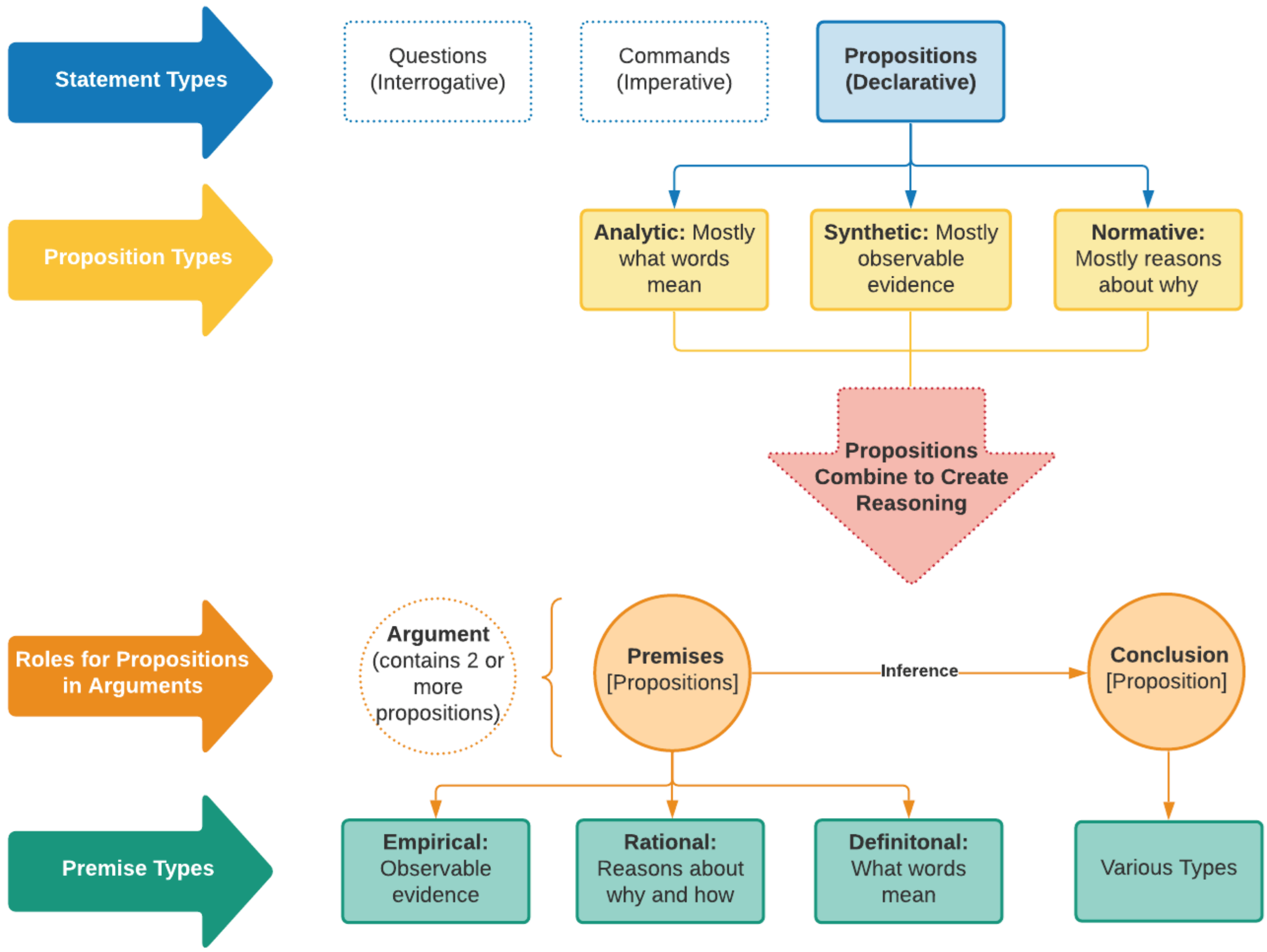


Figure 2.2.1. Summarising statement, proposition, and premise types by Michael Ireland in Mastering Thinking is licensed under CC BY-SA 4.0

Students often struggle to distinguish between empirical and rational propositions when analysing premises in arguments. Simply put, a rational proposition explains the "why" and "how" of something without relying on direct, observable evidence. In contrast, an empirical proposition is grounded in actual observations or data. Evidence, in this context, refers to observable facts or findings that support a proposition. While a reason explains why something might be true, evidence demonstrates that it is true.

For example, consider the debate on the effectiveness of face masks. An argument in favour of enforcing masks might use an empirical premise, citing scientific studies that show communities with enforced mask mandates have lower COVID-19 transmission rates. This is empirical because it relies on observed data and measurable outcomes.

On the other hand, an argument against enforcing face masks might use a rational premise, such as claiming that COVID-19 virus particles are so small that mask fabric cannot block them, likening it to trying to stop mosquitoes with a chicken-wire fence. This reasoning is based on logical inference rather than direct observation. It could, however, be countered with another rational premise, such as the explanation that the virus spreads primarily in larger respiratory droplets, which masks are effective at blocking.

Rational claims often focus on the nature of things rather than specific observations, which can make them seem confusing. For instance, while arguments about the size of COVID-19 particles may appear empirical, they actually rely on reasoning from scientific knowledge about droplet size and mask functionality rather than presenting direct observational data.

At first glance, analysing arguments and their structure might seem overly intellectual or abstract, but this skill is highly practical in everyday life. Every belief, claim, scientific fact, or even marketing message you encounter can be understood as an implicit argument, which is an unstated combination of premises and conclusions. Nearly all beliefs can, and should, be expressed as formal arguments. Reasoning itself is the process of combining propositions into arguments and analysing those arguments to assess the credibility of their claims. Understanding this framework equips you to think critically and evaluate the reliability of the information you encounter.
Every day, we encounter a variety of claims and ideas about the world. For instance:

- Bitcoin values will drop significantly in the next 12 months.
- Systemic racism is a serious and destructive force in modern Western societies.
- Our sports team will win the grand final.
- Human activity is causing global climate warming.
- My political party's candidate is the best choice for president.
- Our new product will cure baldness, insomnia, wrinkles, erectile dysfunction, ageing, and more.
- My birthday party this year will be amazing.
- Critical thinking is a valuable subject to study.

One of the most important critical thinking skills you can develop is the ability to analyse these claims, which are all examples of propositions. This involves breaking them down to understand their argument structure, identifying the premises (reasons) and conclusions, and evaluating whether the premises are true and how well they support the conclusion.

Let us look at a simple example:

Proposition (or claim)	The lecturer is a nerd.

This is a proposition because it makes a claim about something, asserting that it is true. However, on its own, this statement is unconvincing because it does not provide any reasons to support the claim. Without support, it is likely to be dismissed. To justify this proposition, we need to include other statements that provide reasons for believing it. These supporting statements are called premises.

In this case, the proposition "The lecturer is a nerd" becomes the conclusion, and the premises we provide are the reasons that aim to convince others of its truth. Together, these premises and the conclusion form an argument. This process of supporting claims with premises is the foundation of constructing and evaluating arguments effectively (Figure 2.2.2).

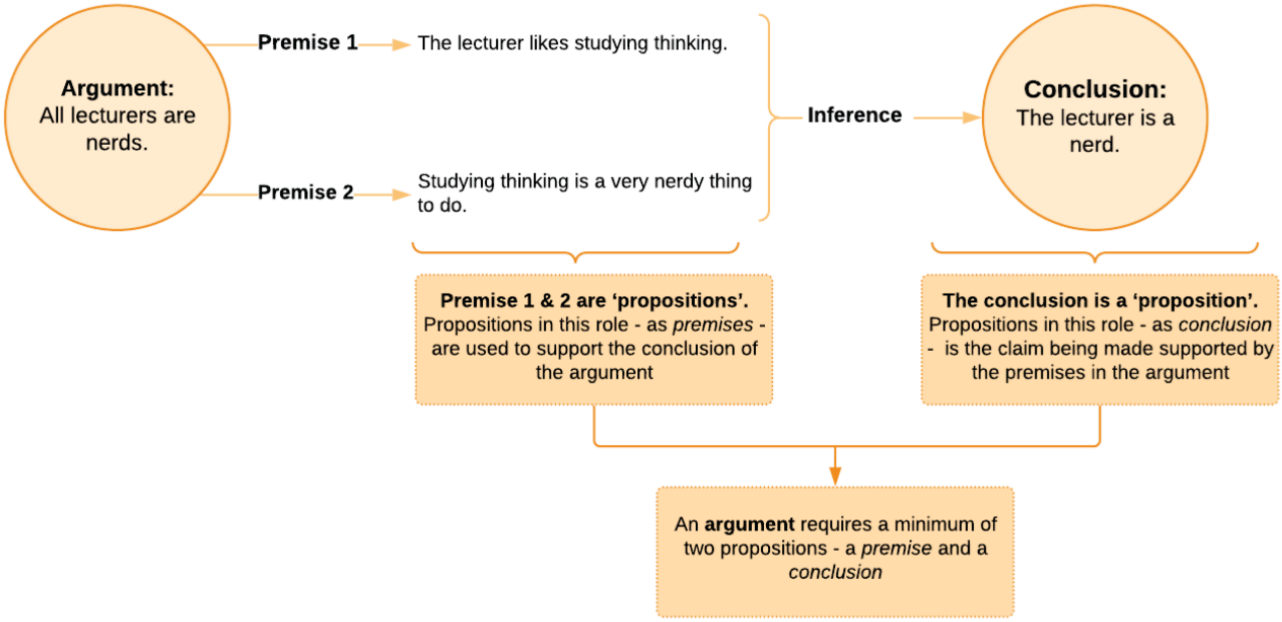


Figure 2.2.2. The structure and pieces of a simple argument by Michael Ireland in Mastering Thinking is licensed under CC BY-SA 4.0

To support my conclusion, we offer two premises. The first is empirical, based on observable evidence. For example, we might notice that the lecturer spends a lot of time studying thinking, or we could ask him, and he confirms this, which serves as evidence. The second premise is rational or definitional, focusing on the nature of studying and claiming it is a "nerdy" activity. This idea cannot be directly observed but is based on reasoning about what it means to be "nerdy".

Interestingly, the premises in this argument could also act as conclusions in other arguments with their own supporting premises. They might even be used to justify a completely different conclusion. However, when evaluating any argument, we must examine it as a whole, considering the relationships between premises and conclusions.

One key reason for learning critical thinking is that our brains do not naturally function optimally. This is the first lesson in critical thinking: our minds are often flawed, prone to deception, and easily misled. This recognition requires humility, which is the willingness to acknowledge that we overestimate our reasoning skills and underestimate our vulnerability to biases and distortions. Because of this, we should be modest about our beliefs and remain open to revising them. Even with strong critical thinking skills, our ideas should always be tentative, as being dogmatic or overly confident often indicates a failure to apply critical thinking. Pride in our opinions is a red flag that we may not be reasoning effectively.

Our brains evolved over millennia, not to find the "truth" but to create models of the world that support survival. Some of the errors in our thinking are deeply ingrained because they served a survival purpose for our ancestors. For example, certain delusions that enhanced survival became hardwired into our brains and are now passed on, often held passionately and without question. While our brains are powerful tools for thinking and survival, they can also limit us by locking us into beliefs that favour survival over truth.

Another reason to learn critical thinking is the environment we live in. Modern life is saturated with individuals and organisations trying to influence our thoughts and beliefs for their own ideological, political, or financial gain. We exist in a highly competitive information landscape, where ideas clash, misinformation spreads, and everyone is vying for our attention. This creates a perfect storm: our natural cognitive flaws combined with an overwhelming flood of marketing, propaganda, clickbait, and biased news make it difficult to discern truth from manipulation.

The good news is that we can improve. Through learning and practice, we can train ourselves to think more clearly and critically, reducing errors, biases, and distortions in our reasoning. These skills not only help us improve our own lives but also enable us to contribute meaningfully to our families, communities, and democratic societies.

Most claims can be expressed as arguments with premises and a conclusion. An argument is a group of statements, where one (the conclusion) is supported by the others (the premises). To illustrate this process, let us consider a simple argument with three premises and one conclusion (Figure 2.2.3).

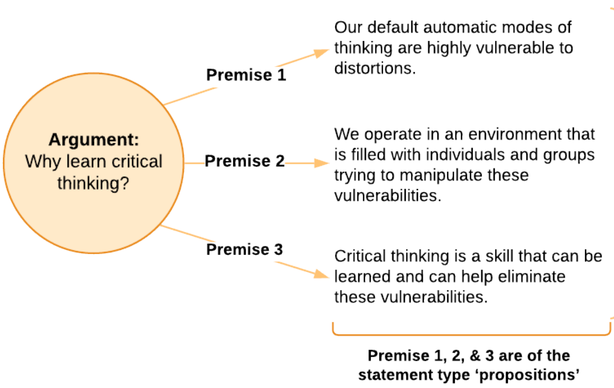


Figure 2.2.3. The structure of an argument by Michael Ireland in Mastering Thinking is licensed under CC BY-SA 4.0

This argument has four statements, all of which are propositions because they make claims about the world that can be true or false. The first three statements are premises, providing reasons to accept the final statement, which is the conclusion. Whether the conclusion is convincing depends on whether the premises are true and how well they logically support the conclusion. Structuring arguments in this way makes it easier to analyse and evaluate their validity and strength.

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

2.3. IMPLICIT PREMISES AND NORMATIVE CONCLUSIONS

By Stephanie Gibbons and Justine Kingsbury, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Implicit Premises

When people present arguments in everyday language, they often leave out parts of their arguments. This is usually because some points are so obvious that they can be safely assumed and do not need to be explicitly stated.

Consider the following argument:

Premise:	My pet Squeaky is a mouse, and all rodents have teeth that never stop growing.
Conclusion:	So Squeaky's teeth will never stop growing.

There is an unstated assumption here: that mice are rodents. Without this assumption, the conclusion does not logically follow from the premises.

The problem with leaving premises unstated is that sometimes the unstated premise is not obvious or easily accepted and may even be highly controversial. Therefore, when reconstructing arguments, we make any implicit premises explicit. This allows us to properly assess each premise as true or false.

Using the example above, we can put the argument into standard form:

Premise 1:	My pet Squeaky is a mouse.
Premise 2:	All rodents have teeth that never stop growing.
Conclusion:	Squeaky's teeth will never stop growing.

We then note that the argument is invalid. We can make it valid by adding a premise:

Premise 1:	My pet Squeaky is a mouse.
Premise 2:	All rodents have teeth that never stop growing.
Premise 3:	**All mice are rodents.**
Conclusion:	Squeaky's teeth will never stop growing.

The argument is now valid. This addition is sensible because it is clearly something the arguer intended, even though it was not explicitly stated.

Sometimes an implicit premise is left out because it is so obvious it hardly needs stating. However, sometimes an unstated premise is doing a lot of work in the argument, and this is not evident because it has not been explicitly stated. Sometimes the unstated premise is obviously false or highly controversial. By exposing implicit premises and making them explicit, we are better positioned to assess the argument.

Consider this argument:

Premise:	Eating dessert every day is unhealthy.
Conclusion:	So no one should do it.

An initial reconstruction might look like this:

Premise:	Eating dessert every day carries a risk of harm to one's health.
Conclusion:	No one should eat dessert every day.

What is the missing premise here? To make the argument valid, a connection needs to be made between the risk of harming one's health and what should not be done. We could add an implicit premise like this:

Premise 1:	Eating dessert every day carries a risk of harm to one's health.
Premise 2:	**No one should do anything that carries a risk to one's health.**
Conclusion:	No one should eat dessert every day.

This is the minimum required to make the argument valid. The arguer must have something like this in mind; otherwise, the conclusion would not follow. Here, the connecting premise (Premise 2) is doing a lot of work in the argument, and it is false. It cannot be true that no one should do anything that would put one's health at risk. If that were true, people would never eat fried food, drink coffee, or participate in sports. Living a life free of risk would be paralysing for anyone.

It is likely that the arguer meant that the risk of eating dessert every day is an unacceptable risk. However, since their argument does not evaluate risk or explain what degree of risk is acceptable, adjusting their argument in this way would be doing too much work for them. In the absence of any attempt to explicitly link the premise to the conclusion, we can only provide the minimally necessary connection and assess it.

Determining what premise needs to be added to an argument to make it valid is tricky. You need to understand how validity works and how to connect the provided information to ensure that the conclusion follows. The following video by Stephanie Gibbons offers some hints to get you started [6:40]. It is a good idea to watch it before attempting the questions.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#oembed-1

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#h5p-4

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#h5p-5

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#h5p-6

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#h5p-7

Normative Conclusions

Arguments with normative conclusions require special attention. They are quite common and often contain implicit premises that need to be made explicit.

A normative claim (or conclusion) is a statement about what should or ought to happen, whereas a descriptive claim (or conclusion) is a statement about how things are. For example, "Mount Everest is the highest mountain in the world" is a descriptive statement, as it describes a current feature of the world. In contrast, "Climbers should seek permission before climbing Mount Everest" and "Fewer people should climb Mount Everest" are normative statements: it goes beyond merely describing how things are and instead prescribes how things should be.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#h5p-8

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#h5p-9

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=836#h5p-10

Conclusions are frequently normative because arguments often aim to persuade people about how things ought to be or what ought to happen.

For an argument with a normative conclusion to be valid, it must include at least one normative premise. No valid argument can consist solely of premises that describe the way the world is and then conclude something about how things should be.

Consider this argument:

Premise:	Some people are finishing their schooling unable to read.
Conclusion:	We should implement a more comprehensive literacy program in our schools.

The conclusion here is normative. The factual claim that some people are unable to read is not sufficient to support the normative claim. No amount of information about how things are can, on its own, support the claim that things should be different. Therefore, we can see that a normative premise is needed for validity.

The argument can be made valid by adding a conditional premise:

Premise 1:	Some people are finishing their schooling unable to read.
Premise 2:	**If some people are finishing their schooling unable to read, then we should implement a more comprehensive literacy program in our schools.**
Conclusion:	We should implement a more comprehensive literacy program in our schools.

Inferring a normative conclusion from descriptive premises is a common type of argument failure known as "the fallacy of deriving ought from is". However, this can generally be avoided by adding the necessary normative premise.

Chapter Attribution

Content adapted, with editorial changes, from:

How to think critically (2024) by Stephanie Gibbons and Justine Kingsbury, University of Waikato, is used under a CC BY-NC licence.

2.4. DEDUCTIVE ARGUMENTS

By Stephanie Gibbons and Justine Kingsbury, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

A deductive argument is one that is intended to guarantee the truth of its conclusion. The terms we use in evaluating deductive arguments are validity/invalidity and soundness/unsoundness.

First, let us discuss validity. A valid argument is one in which, if all its premises were true, the conclusion would also have to be true. For validity, it does not matter whether the premises are actually true. What matters is that there is a connection between the premises and the conclusion such that if the premises were true, the conclusion would necessarily follow. A valid argument is one where it is impossible for the premises to be true and the conclusion false.

The validity of an argument is *independent* of the actual truth of the premises. Therefore, you do not need to know anything about the subject matter of the argument to judge its validity. To say that an argument is valid is to comment on its *structure*, not its content. When we talk about validity, we are addressing the first of the two argument evaluation tasks: the connection between the premises and the conclusion.

A valid deductive argument is one in which, if all the premises were true, the conclusion would also have to be true.

For example:

Premise 1:	No men are mothers.
Premise 2:	Some students are men.
Conclusion:	Some students are not mothers.

and

Premise 1:	All cats can fly.
Premise 2:	Whiskers is a cat.
Conclusion:	Whiskers can fly.

Remember that when considering an argument's validity, it does not matter whether the premises are actually true. So, it does not matter, for the moment, whether it is true that no men are mothers or that all cats can fly. What matters is the connection between the premises and the conclusion. A valid argument has the strongest possible connection between premises and conclusion, so strong that if the premises were all true, the truth of the conclusion would be guaranteed.

In the first example, to see why the argument is valid, think: suppose it is true that no men are mothers and that some students are men. Then, must it also be true (on that supposition) that some students are not mothers?

The answer is that, supposing those premises to be true, it must also be true that some students are not mothers. So, the argument is valid.

Applying the same reasoning to the Whiskers example: Suppose it were true that all cats can fly, and that Whiskers is a cat. Then the conclusion would also have to be true: Whiskers can fly. The argument is valid.

You may be thinking, "But that's absurd! We all know that it is not true that all cats can fly! So how can the argument be valid?"

Bear in mind that validity is not the only consideration in determining whether an argument is good. It also matters whether the premises are true. The Whiskers argument is valid, but it is still not a good argument. We will address this issue shortly.

In everyday language, the word 'valid' is often used to mean 'true' or 'reasonable'. In philosophy and in this book, 'valid' has a technical meaning. A valid argument is one where it is impossible for the premises to all be true and the conclusion false.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-13

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-14

Here are several ways to define a valid argument. They all convey the same concept, so you can use whichever helps you best understand validity.

A valid argument is one in which:

- It is impossible for all the premises to be true and the conclusion to be false at the same time.
- The conclusion logically follows from the premises.
- If all the premises were true, the truth of the conclusion would be guaranteed.
- If all the premises were true, the conclusion would also have to be true.

When assessing whether an argument is valid, you are evaluating its structure, not its content. Consider the following argument:

Premise 1:	All adlers are bobkins.
Premise 2:	All bobkins are crockers.
Conclusion:	All adlers are crockers.

You can determine that this argument is valid even if you do not know what adlers, bobkins, or crockers are. The validity of the argument is based on its *structure*: if the premises were true, the conclusion would necessarily follow.

Typically, you cannot determine an argument's validity or invalidity solely from the truth or falsity of its premises and conclusion. A valid argument can have false premises and a false conclusion, false premises and a true conclusion, or true premises and a true conclusion. An invalid argument can also have these combinations. The only scenario where you can determine an argument's validity from the truth or falsity of its premises and conclusion is when an argument has true premises and a false conclusion. A deductive argument with true premises and a false conclusion must be invalid, as a valid argument can never have true premises and a false conclusion.

Therefore, when deciding whether an argument is valid, do not focus on the actual truth of the premises and conclusion. Instead, imagine or suppose that the premises are true, and then consider whether this would mean that the conclusion also has to be true.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-15

An interactive H5P element has been excluded from this version of the text. You can view it

online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-16

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-17

You might be wondering why we should care about validity. Since validity does not concern the truth of the premises, what is its significance? We should not accept the conclusion of an argument if its premises are obviously false, as in the Whiskers the cat argument or the "sisters and brothers" argument mentioned earlier. So why point out that the argument is valid if it is clearly flawed?

In such cases, you might not need to in real-life contexts. A deductive argument must satisfy two conditions to be considered sound: it must be valid, and it must have true premises. Once we identify that an argument has false premises, we can already determine it is unsound, regardless of its validity.

However, it is crucial to assess the validity of an argument in other situations. One important scenario is when a deductive argument has all true premises. This alone does not make it a good argument; you must also check its validity. For example, consider the following argument:

Premise 1:	February is the next month after January.
Premise 2:	Grass is green.
Conclusion:	Snow is white.

Although the premises are true, they do not logically connect to the conclusion. Therefore, they provide no reason to believe the conclusion. This demonstrates that having true premises is not sufficient to make an argument sound.

Another scenario where assessing validity is important is when others disagree with you about the truth of the premises. In some contexts, it is useful to point out that even if you think the premises are false, the argument would still be invalid if the premises were true.

Common Argument Patterns

Some argument types are so common that they have their own names. Learning to recognise these patterns will help you identify valid arguments.

This section introduces four common argument patterns and some simple variations on them. All of the argument patterns in this section are valid:

- Modus ponens (affirming the antecedent)
- Modus tollens (denying the consequent)
- Disjunctive syllogism
- Hypothetical syllogism
- Some notes on conditionals and generalisations

Modus Ponens (Affirming the Antecedent)

Consider the following argument:

Premise 1:	If Rover is a dog, then Rover is a mammal.
Premise 2:	Rover is a dog.
Conclusion:	Rover is a mammal.

This argument follows the pattern:

Premise 1:	If p then q
Premise 2:	p
Conclusion:	Therefore q

The letters ‘p’, ‘q’, ‘r’, etc., are traditionally used to represent statements. Any statement can be substituted for ‘p’ and ‘q’, and the resulting argument will be valid. Recognising common argument forms makes it easier to identify valid (and invalid) arguments.

Here is another argument with the same pattern:

Premise 1:	If Winston Peters is a Cabinet Minister, then New Zealand First must have entered into a coalition agreement with either the National Party or the Labour Party.
Premise 2:	Winston Peters is a Cabinet Minister.
Conclusion:	New Zealand First must have entered into a coalition agreement with either the National Party or the Labour Party.

Although this argument has longer statements, the basic pattern remains the same. This pattern is known as Modus Ponens, and it is valid.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-18

Modus Tollens (Denying the Consequent)

Modus Tollens is another common valid argument form. It follows this pattern:

Premise 1:	If p then q
Premise 2:	Not q
Conclusion:	Therefore not p

The first premise states that if p occurs, then q must also occur. The second premise points out that q has not occurred. Therefore, it must follow that p has not occurred either. If p had occurred, then q would also have occurred, but we know q has not occurred.

For example:

Premise 1:	If Kamala Harris had won the last election, she would be president.
Premise 2:	Kamala Harris is not president.
Conclusion:	Kamala Harris did not win the last election.

It is clear that this conclusion cannot be false while the premises are true. This is a valid argument.

Try this one. Remember, you are looking to see whether the pattern of Modus Tollens applies.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-19

Try another one:

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-20

Disjunctive Syllogism

A 'disjunction' is a complex statement where two statements are joined with 'or' (or another word serving the same role). A disjunctive syllogism is a valid argument with the following form:

Premise 1:	p or q
Premise 2:	Not p
Conclusion:	Therefore q

This argument form is valid because the initial premise dictates that one of the two options must hold, and the second premise asserts that one does not hold. It follows that the other must hold.

For example:

Premise 1:	Chess is the most challenging board game or Monopoly is the most challenging board game.
Premise 2:	Chess is not the most challenging board game.
Conclusion:	Monopoly is the most challenging board game.

It does not matter whether it is what is before the 'or' or what is after the 'or' that is denied. But it must be denied.

You can practice applying the pattern in the following questions.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-21

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-22

A disjunctive syllogism is often expressed using an 'either… or…' construction. For instance:

Premise 1:	Either there will be a recession, or house prices will continue to rise.
Premise 2:	House prices will not continue to rise.
Conclusion:	There will be a recession.

Sometimes a disjunctive syllogism uses 'either' along with 'or', and sometimes it does not. This does not change the force of 'or' either way. 'Either' is generally used rhetorically to emphasise the contrast between the two options.

Hypothetical Syllogism

A hypothetical syllogism creates a 'chain' of conditional claims. As long as the links of the chain occur in the right way, where each leads to the next, the intermediate links can be omitted. For example:

Premise 1:	If housing prices continue to rise, then rents will continue to rise.
Premise 2:	If rents continue to rise, then rental accommodation will become unaffordable for the working poor.
Conclusion:	If housing prices continue to rise, then rental accommodation will become unaffordable for the working poor.

This follows the general form:

Premise 1:	If p then q
Premise 2:	If q then r
Conclusion:	Therefore, if p then r

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-23

When checking an argument form, the order of the premises is irrelevant. Validity treats the premises as a collection of claims. You can change the order of the premises if it makes it easier for you.

Conditionals and Generalisations

Conditionals

Several of the basic argument patterns above use conditional claims. A conditional is an 'if… then…' statement. The 'if…' part of the statement is called the 'antecedent', and the 'then…' part is called the 'consequent'.

Conditionals are sometimes expressed in a different order. The 'antecedent' is the 'if…' clause no matter what order the parts are presented in. For example, "If Borka is a goose, then Borka is a bird" means the same as "Borka is a bird if she's a goose".

One version of a conditional that people find especially tricky is 'only if'. For example, a sign in a university car park that says, "Staff permit holders only" means "You can only park here if you are a staff permit holder". This does not mean that if you are a staff permit holder you must park there. It means that if you are not a staff permit holder, you must not park there. This is equivalent to "If you park here, then you are (must be) a staff permit holder".

The order of antecedent and consequent in a conditional statement is very important and cannot simply be reversed. For example, "If Borka is a goose, then Borka is a bird" is true, but "If Borka is a bird, then she is a goose" is false because some birds are not geese. The 'only if' claim equivalent to "If Borka is a goose, then she is a bird" is "Borka is a goose only if she is a bird".

An interactive H5P element has been excluded from this version of the text. You can view it

online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-24

Conditional statements can also be expressed using 'unless'. "If Borka is a goose, then she is a bird" is equivalent to "Borka is not a goose unless she is a bird". We often use 'unless' in contrast to a 'not' claim. For example, "I won't babysit for you unless you pay me" means the same as "If I babysit for you, then you are (must be) paying me".

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-25

Conditionals and Generalisations

There is also an important relationship between conditionals and generalisations. The reason why "If Borka is a goose, then Borka is a bird" is true is because the generalisation "All geese are birds" is true.

Any hard generalisation can be expressed as a conditional. "All geese are birds" can be expressed as "If something is a goose, then it is a bird". This means that the basic argument patterns that use conditionals all have forms that use generalisations instead.

Premise 1:	All As are Bs
Premise 2:	x is an A
Conclusion:	Therefore, x is a B

This is a variation on Modus Ponens, using a generalisation in place of a conditional.

Here is an example of a hypothetical syllogism using generalisations instead of conditionals:

Premise 1:	All squares are quadrilaterals.
Premise 2:	All quadrilaterals are polygons.
Conclusion:	All squares are polygons.

Any argument with this form will be valid.

Here are some for you to try:

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-26

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-27

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-28

Basic Structural Fallacies (Formal Fallacies)

A fallacy represents a flawed argument. These fallacies often resemble valid argument patterns and are frequently mistaken for them. However, these forms are invalid. It is essential to practice recognising the differences, paying close attention to the patterns.

This section examines three common structural fallacies:

- the fallacy of affirming the consequent
- the fallacy of denying the antecedent
- disjunctive fallacies.

The Fallacy of Affirming the Consequent

The fallacy of affirming the consequent is an invalid argument often confused with Modus Ponens or Modus Tollens. It follows this structure:

Premise 1:	If p then q
Premise 2:	q
Conclusion:	Therefore p

In this fallacy, the consequent of the first premise is affirmed in the second premise. Such an argument is invalid. The first premise asserts that if p occurs, q must also occur. However, it does not claim that the occurrence of q guarantees the occurrence of p.

Consider the following examples:

Example 1:

Premise 1:	If Bernie Sanders is the president of the US, then Kamala Harris is not the president of the US.
Premise 2:	Kamala Harris is not the president of the US.
Conclusion:	Bernie Sanders is the president of the US.

While the first premise is true (only one person can be president at a time), and the second premise is also true (Kamala Harris is not the president), the conclusion does not logically follow. The absence of Kamala Harris as president does not necessarily mean Bernie Sanders is the president.

Example 2:

Premise 1:	If this shape is a square, then its sides are equal in length.
Premise 2:	This shape's sides are equal in length.
Conclusion:	This shape is a square.

Although any square will have sides of equal length, it is possible for a shape with equal sides to not be a square (e.g., an equilateral triangle). Thus, the conclusion does not follow from the premises.

Recognising the problem in these examples is relatively straightforward. However, some instances of affirming the consequent can be more challenging to identify. It is helpful to carefully analyse the argument's form, often using letters instead of statements to avoid being distracted by known or believed truths.

Can you correctly identify the argument form?

An interactive H5P element has been excluded from this version of the text. You can view it

online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-29

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-30

The Fallacy of Denying the Antecedent

The fallacy of denying the antecedent is another invalid argument often mistaken for Modus Ponens or Modus Tollens. It follows this structure:

Premise 1:	If p then q
Premise 2:	Not p
Conclusion:	Therefore not q

Consider the following example:

Premise 1:	If it is wrong to eat meat, then it is wrong to eat human beings.
Premise 2:	It is not wrong to eat meat.
Conclusion:	It is not wrong to eat human beings.

While the first premise may be true, it does not follow that if eating meat is permissible, eating human beings is also permissible. There can be compelling reasons to avoid eating humans, even if eating other types of meat is acceptable.

Another example:

Premise 1:	If Kamala Harris is president, then the president is a woman.
Premise 2:	Kamala Harris is not president.
Conclusion:	The president is not a woman.

The conclusion does not follow from the premises. It is possible to imagine a scenario where Kamala Harris is not president, but another woman holds the office. Thus, the argument is invalid.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-31

Disjunctive Fallacies

A disjunctive syllogism follows this valid form:

Premise 1:	p or q
Premise 2:	Not p
Conclusion:	Therefore q

However, the following form is invalid:

Premise 1:	p or q
Premise 2:	p
Conclusion:	Therefore not q

There are situations where arguments of the second form might be accepted. Consider the following example:

Premise 1:	Donald Trump or Kamala Harris will win the next presidential election.
Premise 2:	Kamala Harris will win the next presidential election.
Conclusion:	Donald Trump will not win the next presidential election.

At first glance, this argument may appear valid: if both premises are true, the conclusion must also be true. This is because there can only be one winner of the election; it is impossible for both Trump and Harris to win. Thus, if Harris wins, Trump cannot. Here, the true meaning of Premise 1 is "Donald Trump or Kamala Harris will win the next election, but not both".

In English, the word "or" can mean "or, and both are possible" (inclusive or) or "or, but not both" (exclusive or). The valid form of a disjunctive syllogism applies to either interpretation of "or". However, the form:

Premise 1:	p or q
Premise 2:	p
Conclusion:	Therefore not q

is only valid if the "or" is exclusive, meaning it is not possible for both p and q to occur. This determination depends on the meaning of the claim, not merely the argument's form.

When assessing such arguments, ask yourself, "Is it possible for both p and q to occur?" If it is possible, the argument is invalid. If it is impossible, the argument is valid. Adding "but not both" to the disjunctive premise can clarify this.

Sometimes, people assume that an "either… or…" construction indicates an exclusive "or". While this might be useful, it is not always the case in English. For example, if someone says, "Bring either beer or wine to the party; I don't mind", and you bring both, it would be unreasonable for them to say, "You brought both. You can't come in. I said to bring either one or the other". The presence of "either" does not specify whether the "or" is exclusive or inclusive. Use common sense to determine this. A good rule of thumb is to assume the "or" is inclusive if unsure.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=844#h5p-32

Chapter Attribution

Content adapted, with editorial changes, from:

How to think critically (2024) by Stephanie Gibbons and Justine Kingsbury, University of Waikato, is used under a CC BY-NC licence.

2.5. INDUCTIVE ARGUMENTS

By Stephanie Gibbons and Justine Kingsbury, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

The renowned fictional detective Sherlock Holmes often speaks of deducing his conclusions from evidence. However, the reasoning he employs is not actually deductive reasoning.

In this chapter, we use the term "inductive arguments" broadly to refer to all forms of non-deductive reasoning. This includes probabilistic arguments, enumerative inferences, arguments from samples, analogies, causal reasoning, and inference to the best explanation. Although some of these forms are sometimes treated separately in philosophy, here they are all considered types of inductive arguments because they support conclusions without guaranteeing them.

Consider this excerpt from the Sherlock Holmes story "A Scandal in Bohemia":

Dr Watson visits Holmes after a long absence. Holmes figures out that Watson has started practising medicine again, and that he has been out in bad weather lately, and that he has an incompetent servant, even though Watson has not told him any of these things.

> …"my eyes tell me that on the inside of your left shoe, just where the firelight strikes it, the leather is scored by six almost parallel cuts. Obviously they have been caused by someone who has very carelessly scraped round the edges of the sole in order to remove crusted mud from it. Hence, you see, my double deduction that you had been out in vile weather, and that you had a particularly malignant boot-slitting specimen of the London slavery. As to your practice, if a gentleman walks into my rooms smelling of iodoform, with a black mark of nitrate of silver upon his right forefinger, and a bulge in the side of his top hat to show where he has secreted his stethoscope, I must be dull indeed if I do not pronounce him to be an active member of the medical profession."

In this passage, Holmes draws three conclusions:

1. Watson has been out in bad weather.
2. Watson has an incompetent servant.
3. Watson has resumed practising medicine.

The evidence for conclusions 1 and 2 is the leather on the inside of Watson's left shoe, scored by six almost parallel cuts. The evidence for conclusion 3 is the smell of iodoform, a black mark of nitrate of silver on Watson's right forefinger, and a bulge in the side of his top hat.

Even if these are good reasons to believe the conclusions, they do not guarantee them. There are other logically possible reasons for the cuts on Watson's shoe, the smell of iodoform, etc. The unstated premise is that the hypothesis that Watson has resumed practising medicine is the best explanation for the observed facts. Such arguments can never guarantee the truth of their conclusion; it is always possible that another explanation is correct. Therefore, they are not deductive arguments. Nonetheless, non-deductive arguments can provide strong reasons to believe their conclusions. We will now discuss various types of non-deductive arguments, concluding with arguments like Holmes's, known as inference to the best explanation.

One way to indicate that an argument is non-deductive is to precede the conclusion with "Probably". This shows that the conclusion is not guaranteed, only likely. To test whether an argument is deductive or non-deductive, consider whether it makes sense to add "Probably" before the conclusion. For example, if someone argues that since all mice have tails and Minnie is a mouse, Minnie has a tail, it would not make sense to add "Probably" before the conclusion "Minnie has a tail". The premises, if true, make the conclusion certain. Conversely, if the argument is "Almost all mice have tails, and Minnie is a mouse, so Minnie has a tail", it makes sense to add "Probably", indicating a non-deductive argument.

Non-deductive arguments are never valid and therefore never sound. However, since they do not aim to be, we do not criticise them for being invalid or unsound. Instead, we evaluate non-deductive arguments based on their "strength" and "cogency".

Strength in the evaluation of non-deductive arguments serves a similar role to validity in the evaluation of deductive arguments. A non-deductive argument is considered strong if its premises, assuming they are true, provide substantial reason to believe the conclusion, although they do not guarantee its truth. Like validity, strength pertains to the relationship between the premises and the conclusion and is independent of the actual truth of the premises.

An argument is cogent if it is both strong and has all true premises.

When evaluating non-deductive arguments, we ask the same two questions as we do for deductive arguments:

1. What is the connection between the premises and the conclusion?
2. Are the premises true?

For non-deductive arguments, we address the first question by discussing strength. When we refer to cogency, we are addressing both questions.

There is an important difference between validity and strength. Validity is binary; a deductive argument is either valid or invalid. It cannot be partially valid. Strength, however, is a matter of degree. Some non-deductive arguments provide nearly complete support for the conclusion, while others offer minimal or no support.

Consider the following three arguments:

Argument One:

Premise 1:	96% of politicians are dishonest.
Premise 2:	Winston is a politician.
Conclusion:	Winston is dishonest.

Argument Two:

Premise 1:	75% of politicians are dishonest.
Premise 2:	Winston is a politician.
Conclusion:	Winston is dishonest.

Argument Three:

Premise 1:	Most politicians are dishonest.
Premise 2:	Winston is a politician.
Conclusion:	Winston is dishonest.

The first argument is very strong, the second is less strong, and the third is even less strong. Strength is a matter of degree.

A strong argument is one in which the premises provide significant support (though not conclusive) for the conclusion: if the premises were true, the conclusion would likely be true.

Strength is a matter of degree, unlike validity.

A cogent argument is a strong argument with all true premises.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-33

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-34

Probabilistic Arguments

Probabilistic arguments occur when the likelihood of the conclusion can be clearly established given the premises. These arguments closely resemble deductive arguments in their structure.

Consider this argument:

Premise 1:	All sheep in New Zealand live on farms.
Premise 2:	Alice is a sheep in New Zealand.
Conclusion:	Alice lives on a farm.

Assume for a moment that Alice is a New Zealand sheep (i.e., Premise 2 is true). The argument is valid. However, it cannot be sound because the first premise is a "hard" generalisation, which does not allow for any exceptions. As a hard generalisation about all sheep in New Zealand, Premise 1 is false. There are undoubtedly some rogue sheep, some that have escaped into the bush, and some kept as pets that do not live on farms. Thus, although the argument is valid, it is unsound.

We could modify Premise 1 to a "soft" generalisation, which has a better chance of being true. A soft generalisation makes a general claim about a group but allows for exceptions. For example, if Premise 1 stated "Nearly all sheep in New Zealand live on farms", then it would be true.

However, the argument would no longer be valid:

Premise 1:	Nearly all sheep in New Zealand live on farms.
Premise 2:	Alice is a sheep in New Zealand.
Conclusion:	Alice lives on a farm.

In this argument, the premises do not guarantee the conclusion. It is possible for the premises to be true while the conclusion is false, as Alice could be one of the few rogue bush-sheep or a pet.

This type of argument is not valid, but can be very useful. The premises do not guarantee the conclusion, making the argument invalid. However, it provides strong support for the conclusion. The truth of the premises is sufficient to show that the conclusion is probably true.

This type of argument is not a failed deductive argument; it does not intend for the conclusion to follow with certainty. We can indicate this in the argument frame by including the word "Probably" before the conclusion, as follows:

Premise 1:	Nearly all sheep in New Zealand live on farms.
Premise 2:	Alice is a sheep in New Zealand.
Conclusion:	[Probably] Alice lives on a farm.

Arguments of this nature will vary in strength depending on how probable the premises make the conclusion. Some arguments have premises that make their conclusions very probable and are thus very strong.

Premise 1:	There are 99 black marbles in this bag and one white marble.
Premise 2:	In my fist is a marble randomly selected from the bag.
Conclusion:	[Probably] The marble in my fist is black.

Here, it is 99% probable that the marble in my fist is black, making this a very strong argument.

It is important to note that the statement "The marble in my fist is black" is still either true or false. It cannot be 99% true. It is either 100% true or 100% false. The 99% applies to the probability that it is true, not to the truth itself.

We can change the probabilities in such arguments by altering the proportions of marbles:

Premise 1:	There are 75 black marbles in this bag and 25 white marbles.
Premise 2:	In my fist is a marble randomly selected from the bag.
Conclusion:	[Probably] The marble in my fist is black.

This conclusion is still probable. This non-deductive argument is weaker than the previous one but remains strong enough to be useful.

In the marbles example, it is easy to accurately measure the degree of probability of the conclusion. Most ordinary probabilistic arguments lack this level of precision.

Premise 1:	Most university students do not have children.
Premise 2:	Betty is a university student.
Conclusion:	[Probably] Betty does not have children.

Here, the conclusion is probable, but we cannot assign a precise degree of probability to it.

Argument Patterns

The same types of argument patterns can occur in probabilistic non-deductive arguments as in deductive arguments.

The argument "No mammals lay eggs. Perry is a mammal. Therefore, Perry does not lay eggs" is a valid argument. It follows the general pattern of Modus Tollens. (If this is unclear, try converting the generalisation in the first premise into a conditional: "If something is a mammal, then it does not lay eggs".) However, the first premise of this argument is false. There are three species of mammals that lay eggs, the most well-known being the platypus. Thus, we can soften the generalisation in the first premise:

Premise 1:	Hardly any mammals lay eggs.
Premise 2:	Perry is a mammal.
Conclusion:	[Probably] Perry does not lay eggs.

This follows the Modus Tollens pattern, except it uses a soft generalisation instead of a hard generalisation in the first premise. It is a non-deductively strong argument.

It is important to remember that a fallacious argument pattern cannot be improved by weakening the generalisation. Here is an example to illustrate this point:

Premise 1:	All geese are birds.
Premise 2:	Borka is a bird.
Conclusion:	Borka is a goose.

The basic pattern of this argument is the fallacy of affirming the consequent (using a generalisation instead of a conditional). There are many birds that are not geese, and Borka could be one of those.

This argument cannot be improved by weakening the generalisation in Premise 1. This would result in an argument like this:

Premise 1:	Most geese are birds.
Premise 2:	Borka is a bird.
Conclusion:	[Probably] Borka is a goose.

This is a weak argument. Borka is not likely to be a goose simply because it is a bird. Once again, Borka could be another type of bird. Thus, this argument commits a non-deductive version of the fallacy of affirming the consequent. The fundamental problem with the structure of the argument cannot be resolved by changing one of the premises from a hard generalisation to a soft one.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-35

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-36

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-37

Kinds of Soft Generalisation

Any statement that makes a claim about a group or category of things can be considered a generalisation. Hard generalisations include terms like "All", "None", "Always", and "Never". Soft generalisations include terms such as "Almost all", "Almost none", "Many", "Most", and "Some". Some soft generalisations are useful in probabilistic arguments, while others are not.

The goal is to demonstrate that the conclusion is probable. To have any strength, the argument must show that the conclusion is more likely to be true than not.

Consider this argument:

Premise 1:	The majority of people enjoy ice cream.
Premise 2:	Alex is a person.
Conclusion:	[Probably] Alex enjoys ice cream.

This argument has some strength, although not much. If the premises were true, the conclusion would be more likely to be true than not. However, Premise 1 might not be accurate, so the argument cannot be cogent.

We might attempt to improve the argument by further softening the generalisation in Premise 1 to make it more likely to be true. This might result in:

Premise 1:	Many people enjoy ice cream.
Premise 2:	Alex is a person.
Conclusion:	[Probably] Alex enjoys ice cream.

While Premise 1 is now more likely to be true, the argument remains weak. The premises do not provide a strong reason for accepting the conclusion. The term "many" does not specify a proportion of all people that could make the conclusion probable. Words like "many" do not convey precise proportions; they merely indicate that there are at least several people who do so. It is important to consider whether the generalisation is sufficiently robust to make the conclusion probable.

Extended Probabilistic Arguments

Just as deductive arguments can be extended, so can non-deductive ones. It is important to consider how any probability within the argument affects the probability of the final conclusion.

In an extended argument with a single soft generalisation, the probability of the conclusion will reflect the degree of probability in the soft generalisation:

Premise 1:	Nearly all university students write assignments on computers.
Premise 2:	Betty is a university student.
Conclusion 1:	[Probably] Betty writes her university assignments on a computer.
Premise 3:	Everyone who writes assignments on a computer can read.
Conclusion 2:	[Probably] Betty can read.

This is a strong argument. The probability of Conclusion 2 is the same as that of the intermediate Conclusion 1, and the degree of probability of Conclusion 1 comes from the soft generalisation in Premise 1.

However, each additional soft generalisation in an extended argument will further dilute the probability of the final conclusion.

Consider this argument:

Premise 1:	Most university students hand in their assignments.
Premise 2:	Conrad is a university student.
Conclusion 1:	[Probably] Conrad hands in his assignments.
Premise 3:	Most students who hand in their assignments pass their courses.
Conclusion 2:	[Probably] Conrad passes his courses.

Here, the inference from Premise 1 and Premise 2 to Conclusion 1 is not particularly strong. It is further weakened by the soft generalisation in Premise 3. By the time Conclusion 2 is reached, the probability assigned to the final conclusion by the premises is low. This argument is not strong.

If the dilution issue is not clear from that example, consider this one, where the problem is more evident:

Premise 1:	Most of those currently in the university library are university students.
Premise 2:	Conrad is currently in the university library.
Conclusion 1:	[Probably] Conrad is a university student.
Premise 3:	Most university students drink in the evenings.
Premise 4:	It is evening.
Conclusion 2	[Probably] Conrad is drinking.

It is unlikely (though not impossible) that Conrad is drinking in the library. Even if all the premises of this argument were true, the conclusion is not likely. This is because the group of people likely to be in the university library in the evening is different from those who are likely to be drinking.

Sometimes the generalisations in an extended argument will be strong enough to make the final conclusion probable, and sometimes they will not. There is no precise way to determine the probability of the conclusion when using imprecise quantifiers such as "nearly all" and "few". Instead, consider the number and type of generalisations made and make a judgement call about whether the probability of the conclusion has been diluted too much.

A Note on the Use of "Probably"

When presenting non-deductive arguments in standard form, we often insert "[Probably]" before the conclusion to indicate that the argument is intended to be non-deductive. The square brackets signify that this term is not part of the conclusion itself or the argument itself; it merely indicates the type of argument being used. While this can be helpful, it is important to note that it does not reflect the success of the argument. Additionally, inserting "Probably" cannot improve a weak argument. Consider the following example:

Premise 1:	Nearly all dogs have four legs.
Premise 2:	Fido is a dog.
Conclusion:	Fido has four legs.

This is a strong argument. It remains strong regardless of whether "Probably" is placed before the conclusion.

Furthermore, inserting "Probably" before the conclusion does not indicate that an argument is strong, nor will it improve a weak argument. Consider this example:

Premise 1:	Nearly all dogs have four legs.
Premise 2:	Fido has four legs.
Conclusion:	[Probably] Fido is a dog.

This is not a strong argument, as it commits the fallacy of affirming the consequent. The presence of "Probably" cannot change this. You should view "Probably" as a useful way to indicate that an argument is non-deductive, but it does not provide any information about the argument's success.

Enumerative Inferences

Imagine a turkey living happily on a turkey farm. Each morning, the farmer brings corn for it to eat, which is enough to keep the turkey happy. One morning, as the farmer approaches, the turkey might think, "Hooray, breakfast!". If the turkey is reasoning at all, it is reasoning non-deductively: every morning so far, the farmer has brought corn, so today he will bring corn again. Unfortunately, it is Christmas morning, and the turkey makes a grave mistake by running happily towards the farmer, who this time is carrying an axe.

The turkey's reasoning, if it was reasoning at all, was quite sound: it was based on true premises that provided substantial reason to believe its conclusion. Nevertheless, the conclusion was false. The moral is that no matter how strong your non-deductive argument, it is still possible for the conclusion to be false.

Someone reasoning in this manner is taking a large number of observed cases and inferring that a pattern will continue. They have collected data and are extrapolating from it to formulate a conclusion. Inferences of this type are sometimes called "inductive inferences". However, since this is not the only type of induction, we refer to them as "enumerative inferences". This term reflects the process of collecting a number of cases and reaching a conclusion about a new case based on that list.

Enumerative inferences differ from probabilistic arguments. Consider this probabilistic argument:

Premise 1:	There are 75 black marbles in this bag and 25 white marbles.
Premise 2:	In my fist is a marble randomly selected from the bag.
Conclusion:	[Probably] The marble in my fist is black.

In this argument, the proportions of the contents of the bag are known, and because this is a mathematical example, the degree of probability of the conclusion can be precisely calculated (it is 75% likely that the marble in my fist is black).

Now, suppose I have a bag of 100 marbles, but I know nothing about their colours. I draw out the first 99 marbles, and they are all black. Based on this, I conclude that the 100th marble will also be black. My argument looks like this:

Premise 1:	Marble 1 is black.
Premise 2:	Marble 2 is black.
Premise 3:	Marble 3 is black.
	⋮
Premise 99:	Marble 99 is black.
Conclusion:	[Probably] Marble 100 is black.

I cannot assign a precise degree of probability to this conclusion. There are infinite possibilities for the colour of the remaining marble. However, it seems more reasonable to suppose it is black, given the contents of the bag so far, than to suppose it is another colour.

In everyday life, we reason like this frequently. When I assume the sun will rise tomorrow, I am extrapolating from many instances of the sun rising. This has happened every day of my life, and I expect it to continue. Similarly, I assume that if I get hit by a bus, I will be injured, based on what usually happens when people are hit by buses and my past experiences with large, heavy objects. Even the belief that the laws of physics will continue to apply is justified through an enumerative inference. Such arguments are very important and useful.

Not all enumerative inferences are strong, and they can be difficult to assess. Consider the marbles example again. When I know there are 100 marbles in the bag and the first 99 are black, it seems reasonable to conclude that the 100th marble will be black. But what if I did not know how many marbles were in the bag? What if I had only drawn 10 marbles? Can I still justifiably conclude that the next marble will be black?

Several factors must be considered when assessing an enumerative inference:

- sample size

The more data collected, the stronger the enumerative inference. This is why an inference about the colour of the next marble is stronger when 99 marbles have been tested compared to only 9. The sample size of sunrises is enormous, making us very confident that the sun will rise tomorrow.

- sample size relative to the total population

If I know there are a million marbles in the bag and I have tested 99, I will feel less confident about the next marble than if there are 100 marbles in the bag and I have tested 99.

The size of the total population can also vary depending on the conclusion. Sometimes a conclusion is about the next case alone. Consider:

Argument 1:

Premise 1:	The sun has risen every day of my life.
Conclusion:	[Probably] The sun will rise tomorrow.

I feel very confident about this conclusion. The sample size is all the days of my life, and the total population is all the days of my life plus one (i.e., the next day). The sample is a large proportion of the total population.

Compare:

Argument 2:

Premise 1:	The sun has risen every day of my life.
Conclusion:	[Probably] The sun will rise every day for the rest of my life.

and

Argument 3:

Premise 1:	The sun has risen every day of my life.
Conclusion:	[Probably] The sun will rise every day forever.

In Argument 2, the total population is unknown, but I optimistically assume I am halfway through my life. This means I am extrapolating from known cases to about the same number of future cases.

In Argument 3, the conclusion is so broad that it is unlikely to be true. We know the world will end someday, so there will eventually be a last day. A conclusion that extends too far beyond its sample results in a sample that is a very small proportion of the total population.

- sample collection method

Suppose I am given 10 bags of marbles, each containing 10 marbles. If I take 10 marbles from the first bag and they are all black, I have some reason to think all the marbles in all the bags are black, but it is not a particularly strong reason. If I take one marble from each bag and they are all black, I have a much stronger reason to think all the marbles are black.

Generally, the more data collected and the more random the data collection, the stronger the enumerative inference. If you find yourself rejecting almost all enumerative inferences, your standard for reasonableness is likely too high. We use these inferences all the time, and it would be impossible to function without them. You would have no reason not to step in front of a bus.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-38

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-39

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-40

It is important to note that the possibility of being wrong is not sufficient grounds for rejecting an enumerative inference. This possibility of inferring a false conclusion from true premises is a feature of all non-deductive arguments. Consider the turkey example again. The turkey is justified in its conclusion, even though it will eventually be wrong. One day the sun will not rise, but that does not mean you are unjustified in believing it will rise tomorrow.

Arguments from Samples

When we sample or survey some (but not all) members of a group and then draw a conclusion about the group as a whole, we are engaging in non-deductive reasoning. Consider the following example:

A nationwide poll of a random sample of thousands of homeowners revealed that 70% of them oppose increases in social welfare payments. Therefore, it was concluded that roughly 70% of the adult population of New Zealand opposes such increases.

In this case, there was an evident issue with the argument: all the individuals surveyed were homeowners, yet the conclusion was drawn about the entire adult population, not just homeowners. The sample was not representative.

Now, suppose we conduct a more accurate survey: instead of only asking homeowners, we draw our sample randomly from the adult population of New Zealand. Suppose the results indicate that 55% of the sample of thousands of adult New Zealanders oppose increases in social welfare payments. We then conclude that 55% of all adult New Zealanders oppose such increases.

This is a stronger argument than the previous one because the sample from which we are generalising is representative (as far as we can tell) of the wider population. However, it is important to note that the argument remains non-deductive. Unless every single member of the wider population is polled (in which case it is no longer an argument from a sample), the conclusion that what is true of the sample is also true of the wider population is not guaranteed.

In addition to the representativeness of the sample, we must also consider the size of the sample. In the example above, if we had only surveyed 10 randomly selected New Zealand adults instead of thousands, we should not generalise the results to the entire adult population of New Zealand as the sample size would be much too small.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-41

Analogy

An analogy highlights the similarities between two different things.

For example, if I say, “The mind is like a computer: it takes certain inputs, processes them, and then produces results", I am drawing an analogy. I am suggesting that the mind and a computer are alike in some way. However, I am not presenting an argument or drawing any conclusion from this analogy. I am merely drawing parallels, perhaps to encourage you to think about the mind differently or to illustrate how the mind functions.

An argument by analogy, on the other hand, involves pointing out the similarities between two or more things and then drawing a conclusion based on those similarities.

Consider a scenario where I am deciding whether to purchase a particular car. It is a ten-year-old Honda Civic with 75,000 kilometres on the odometer, a little rust, and only one previous owner who drove it carefully and serviced it regularly.

I might reason as follows:

My previous car was a Honda Civic, and when I bought it, it was ten years old with 75,000 kilometres on the odometer, a little rust, and only one previous owner who had maintained it well. That car served me well for five years, requiring minimal repairs. Therefore, this car, being similar to my last one, will likely serve me well too, and I should buy it.

Premise 1:	Car A was a Honda Civic, ten years old, with 75,000 kilometres, a little rust, well-maintained, and it served me well for five years with minimal repairs.
Premise 2:	Car B is a Honda Civic, ten years old, with 75,000 kilometres, a little rust, well-maintained.
Conclusion:	Car B will last five years and require minimal repairs.

General Structure of Arguments by Analogy:

Premise 1:	A has characteristics W, X, Y, and Z.
Premise 2:	B has characteristics W, X, and Y.
Conclusion:	B will also have characteristic Z.

Considerations for Evaluating an Argument by Analogy:

- How similar are the things being compared?
- Are the similarities relevant? (For example, if the similarities mentioned were all about colour, which is irrelevant to the car's performance, it would not be a strong argument.)
- Are there any relevant differences between the things being compared?
- How many similar cases are we dealing with? (For instance, if I had owned three cars with similar characteristics that all served me well, the argument would be stronger.)

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-42

Causal Reasoning

Suppose several individuals experience upset stomachs after a dinner party. Here are the details of what the various attendees consumed:

Foods eaten by those who became ill:

- person A: ham, potato salad, coleslaw
- person B: ham, rice salad, lettuce salad
- person C: ham, pasta salad, carrot salad.

Foods eaten by those who did not become ill:

- person D: chicken, rice salad, coleslaw
- person E: sausages, pasta salad, lettuce salad
- person F: bean salad, potato salad, carrot salad.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-43

It is likely that the ham caused the illness.

Why?

- All those who became ill consumed ham.
- All those who did not become ill did not consume ham.
- There is no other food item that was exclusively consumed by those who became ill.
- Consuming spoiled ham is a known cause of upset stomachs, and we understand the mechanism by which this occurs (unlike, for example, a shared characteristic such as wearing red shirts).

It is important to note that the ham could have caused the illness even if not all ham-eaters became ill. Consuming ham might have increased the probability of illness without guaranteeing it, as some individuals may have stronger constitutions.

Causal statements assert that one thing causes or does not cause another. For example, smoking causes lung cancer, drinking coffee after dinner keeps me awake, and reading logic textbooks after dinner makes me sleepy.

Causal statements are common in both everyday conversation and scientific research. Understanding the effects of actions is crucial for decision-making. Doctors need to know the causes of diseases to treat them effectively, and airlines need to determine the causes of plane crashes to prevent future incidents.

Causal *arguments* consist of a causal claim supported by reasons for believing that claim. For instance, if American Airlines claims that a plane crash occurred because the altimeter malfunctioned and visibility was poor due to low clouds, their reasons might include records showing the altimeter reading fifteen thousand feet just before the crash, despite the mountain being much shorter, and a tape recording of the pilot's exclamation upon seeing the mountain emerge from the fog. Listing these reasons as premises and the causal claim as the conclusion forms a causal argument.

Causal arguments are non-deductive. In the plane crash example, even with all the evidence, it is not 100% certain what caused the crash. However, it can still be a strong argument.

Consider a more general causal claim: Attending St Peter's Cambridge leads to better NCEA results. Suppose statistical analysis shows that the average marks of St Peter's students are higher than the national average. Does this provide good reason to believe the causal claim?

No, not on its own. Correlation does not imply causation. Other possibilities should be considered before accepting such a causal argument:

1. Coincidence: It might be pure chance that St Peter's students performed better.
2. Common cause: There might be an underlying factor that both increases the likelihood of attending St Peter's and achieving good marks, such as having wealthy parents or parents who prioritise education. To rule out these alternatives, a more complex study should be conducted. Compare a group of students similar to St Peter's students in all relevant respects except for the school they attend. If St Peter's students perform better than this control group, and all other relevant factors have been considered, the causal claim is more justified.
3. Opposite direction of causation: Sometimes, the cause and effect are mistaken. For example, New Hebrides Islanders believed lice caused good health because healthy individuals were infested with lice, while sick individuals were not. In reality, lice left their hosts when they developed a fever. Thus, getting sick caused the absence of lice, not the other way around. Observing the order of events can help determine the direction of causation.

Having a theory to explain the causal process is also important. Discovering that lice do not like high temperatures provides additional reason to believe that illness causes the absence of lice rather than the absence of lice causing illness.

Here is a causal argument:

Premise 1:	Most people who take mega-doses of Vitamin C when they have a cold recover within a week.
Conclusion:	Mega-doses of Vitamin C cure colds.

We are justified in believing this conclusion only if we have considered and ruled out likely alternatives.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=851#h5p-44

It might be that people naturally recover from colds within a week, with or without Vitamin C. This can be tested by collecting data on the recovery speed of people who do not take mega-doses of Vitamin C.

Alternatively, those who take Vitamin C might also engage in other health-promoting behaviours, such as eating chicken soup and going to bed early, which could be the actual factors contributing to their quick recovery. To test this, observe a control group that is similar to the test group in all relevant respects (diet, sleeping habits, etc.) except for taking Vitamin C, and compare the two groups.

Inference to the Best Explanation

An inference to the best explanation occurs when you have a phenomenon or observation that requires an explanation, and you conclude that the best available explanation is true simply because it is the most plausible explanation for that phenomenon. Sherlock Holmes' reasoning presented earlier is likely an example of an inference to the best explanation, although he does not explicitly outline all the steps. For instance, he concludes that Watson has resumed practising medicine based on the bulge in Watson's top hat, which would be caused by carrying a stethoscope (implicitly assuming that only practising doctors carry stethoscopes).

To fully articulate the argument, Holmes would need to consider alternative explanations for the bulge in the top hat and justify why the explanation that Watson has resumed practising medicine is the best one. However, it is evident that this is Holmes' line of reasoning, where Watson's return to medicine is the best explanation for the bulge in his top hat, thus providing a reason to believe that Watson has indeed resumed his medical practice. This form of reasoning is common in detective stories and scientific reasoning. Often, the reason we believe in certain unobservable entities (such as electrons) is that their existence provides a good explanation for observable phenomena (e.g., the fact that lights turn on when a switch is flipped).

The argument from design, which argues for the existence of God, can be construed as an inference to the best explanation. It can be outlined as follows:

Observations:

- Organisms are complex and intricate.
- They are well adapted to their surroundings.
- Their parts work together to enable the whole organism to function.

Possible Explanations:

1. God designed organisms to be just the way they are.
2. Organisms evolved by natural selection without any supernatural involvement.
3. Organisms evolved by natural selection, but God designed them to do so.
4. God created organisms 6,000 years ago in such a way that it appears they have been around much longer and have evolved.
5. Organisms came to be the way they are through completely random processes.

Explanation 5 is not a very good explanation. When evaluating which explanation is the best, the first consideration is whether the observations would be surprising if the explanation were true, or if they would be expected. Typically, we seek explanations for surprising phenomena. A good explanation makes the phenomenon unsurprising. Explanation 5 fails this test as it leaves the complexity and intricacy of organisms unexplained.

The other four explanations pass this initial test. To complete the argument, since the conclusion is that God exists, we need reasons to believe that either 1, 3, or 4 is a better explanation than 2. There is evidence against 1 (such as fossils and vestigial organs). However, there is no scientific evidence that decisively distinguishes between 2, 3, and 4. One reason for preferring 2 might be its simplicity. One reason for preferring 3 might be that it explains more, such as the origins of life, which 2 does not. (We are not resolving this question here; this is merely an illustration of how inferences to the best explanation work)

Another example: suppose you observe that milkmaids do not contract smallpox even when smallpox is widespread. The explanation might be that milkmaids contract cowpox, a relatively mild illness, which provides immunity to smallpox.

The fact that milkmaids do not contract smallpox does not conclusively prove that cowpox provides immunity to smallpox. There could be other explanations. Perhaps cows have magical properties that protect those who spend time with them from smallpox. Perhaps milkmaids consume more milk, which contains a substance that protects against smallpox.

What makes these alternative explanations less plausible?

- Consistency with other accepted theories: What seems like a good explanation depends partly on your background assumptions. For example, I assume that magic does not operate in the world, so I do not need to appeal to magic to explain everyday phenomena. You might not share this assumption.
- Results of experimental testing: Do other milk drinkers have immunity from smallpox? (As it happens, they do not.)

Inferences from evidence to explanations are not deductively valid. It is always possible that the explanation is incorrect, despite the evidence.

Chapter Attribution

Content adapted, with editorial changes, from:

How to think critically (2024) by Stephanie Gibbons and Justine Kingsbury, University of Waikato, is used under a CC BY-NC licence.

2.6. SOUNDNESS AND COGENCY

By Stephanie Gibbons and Justine Kingsbury, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

We have previously discussed how to evaluate the structure of an argument, which involves examining the connection between the premises and the conclusion. The terms used to describe the structure of arguments are valid and invalid (for deductive arguments) and strong (for non-deductive arguments).

However, a good argument requires more than just validity or strength. A valid or strong argument might have premises that are false, and in such cases, the conclusion should not be accepted based on those premises, despite the argument's validity or strength. A good argument must also have true premises. When evaluating whether an argument is persuasive, consider whether it is sound (for deductive arguments) or cogent (for non-deductive arguments).

Some definitions:

- A sound argument is valid and has all true premises.
- A cogent argument is strong and has all true premises.

We will now discuss how to assess the truth of an argument's premises.

Remember: for an argument to be sound, it must be both valid and have true premises.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=861#h5p-46

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=861#h5p-47

We already know how to assess the validity or strength of an argument. To determine whether an argument is sound or cogent, we must evaluate the truth of its premises.

People often hesitate to determine the truth value of a statement, saying things like "it's just an opinion" or "there isn't really one truth". However, assessing the truth of statements is essential in argument evaluation, and it is important to make an effort to do so.

Is There Such a Thing as Truth?

A statement is true if it accurately describes the world and false if it does not. This common-sense account of truth will be used in this book.

Some statements are uncontroversially true, such as:

- Squares have four sides.

Others are straightforwardly false, like:

- Squares have three sides.

Regardless of how forcefully someone asserts the latter, it remains false.

Some statements are true at certain times and false at others. This variability does not negate their truth or falsity but reflects changes in the world. For example:

- Joe Biden is the President of the United States.

At the time of writing, this statement is true. At some time in the future, it will be false. This temporal aspect does not affect the statement's current truth.

The truth of claims can change, affecting the soundness or cogency of arguments. For instance:

Premise 1:	Joe Biden is the President of the United States.
Premise 2:	Joe Biden is a man.
Conclusion:	The President of the United States is a man.

This argument is sound as long as Joe Biden is President. When he is no longer President, the argument will cease to be sound.

Some statements are obviously true, while others require investigation to establish their truth or falsity.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=861#h5p-45

If you needed to look up the answer to the question, that is acceptable. You are not expected to know everything already.

Sometimes, assessing the truth of a statement is challenging. For example:

- It is raining.

This statement might be true now but false tomorrow, or true in one location but not another. The difficulty lies in the statement's specificity. Generally, when someone says "It is raining", they mean "It is raining here, at the moment". Making statements more specific by including time and place references can help, but it is often unnecessary.

In some cases, it is impossible to determine the truth or falsity of a statement, even after research. When this happens, you may need to suspend judgement on the statement's truth and the argument's soundness or cogency. However, do not give up too soon when assessing truth. If you find yourself frequently saying "who really knows?" you may be giving up prematurely.

Scepticism is a philosophical position that involves doubting the truth of nearly everything. Even sceptics live their lives as if many ordinary claims are true. While you may adopt scepticism, it is important to take a more practical approach to truth when assessing everyday arguments.

Some propositions are more difficult to assess than others. It is important not to give up but to pause and think carefully.

Consider the statement:

- Most of the people reading this textbook are enrolled in a course.

Is this true?

To evaluate its truth, note that most people in the world are not enrolled in any course. If the statement is true, it must be specific to people who read this textbook. Not all readers are enrolled in a course, but many are, especially if the textbook is assigned for a course. The likelihood of the textbook being a bestseller read for fun is low. Therefore, it is reasonable to conclude that most readers are enrolled in a course.

Determining truth is not always easy, but it is possible and important.

Cogent Arguments and False Conclusions

In a deductively valid argument, the truth of the premises guarantees the truth of the conclusion. This means that a valid argument with true premises must have a true conclusion, making sound arguments particularly useful.

In non-deductive arguments, the relationship between the premises and the conclusion is different. In a strong argument, the truth of the premises makes the conclusion likely but does not guarantee it. Consequently, it is possible to have a cogent argument with a false conclusion.

Consider the following argument:

Premise 1:	Nearly all of the presidents of the United States have been white men.
Conclusion:	[Probably] The 44th president of the United States was a white man.

This is a strong argument: the truth of the premise does not guarantee the truth of the conclusion, but it does make it very likely. There is no issue with the form of the argument.

The premise is true (there have been 46 US presidents, and 45 of them have been white men.)

Since we have a strong argument with true premises, this argument is cogent.

However, note that the 44th president of the United States was Barack Obama, who is not a white man. Thus, this is a cogent argument with a false conclusion.

We can make this even more explicit:

Premise 1:	Forty-five of the 46 presidents of the United States have been white men.
Premise 2:	Barack Obama was the 44th president of the United States.
Conclusion:	[Probably] Barack Obama was a white man.

This argument remains cogent.

This is a consequence of the way cogency is defined. Any strong argument with true premises will be cogent, even if the conclusion is false.

This implies an additional step when assessing cogent arguments compared to sound arguments. In a sound argument, the truth of the conclusion is guaranteed and does not need to be independently assessed. In a cogent argument, once the argument has been established as cogent, it is prudent to consider whether there is any additional reason to believe the conclusion is false. While uncommon, there may be additional information that indicates the conclusion is false. This does not negate the cogency of the argument (as cogency is determined by its definition), but it is a relevant consideration in the assessment of cogent arguments.

Chapter Attribution

Content adapted, with editorial changes, from:

How to think critically (2024) by Stephanie Gibbons and Justine Kingsbury, University of Waikato, is used under a CC BY-NC licence.

2.7. SUMMARY

By Marc Chao

Summary

Forming arguments is a fundamental aspect of critical thinking, particularly in psychology and other disciplines, as it enables systematic evaluation of theories, evidence, and complex issues. An argument, in this context, is defined as a set of propositions where premises provide reasons to support a conclusion. Unlike casual disagreements, effective arguments rely on logical reasoning and evidence to persuade. Propositions, the foundation of arguments, are declarative statements that can be true or false. They serve specific roles, with premises supporting conclusions through inference, a logical connection between ideas. For instance, the premises "All humans are mortal" and "Socrates is human" lead to the conclusion "Socrates is mortal". Critical thinking involves constructing such arguments and critically evaluating the validity of premises and their relationship to conclusions, fostering informed and reasoned decision-making.

Propositions, as the building blocks of arguments, assert claims about the world that can be evaluated for truth or falsity. Effective arguments rely on credible premises, often assumed to be self-evident or justified separately, to strengthen their conclusions. Propositions may be empirical, based on observable evidence, or rational, derived from reasoning about concepts. Critical thinking requires analysing arguments by identifying premises and conclusions, evaluating their truth, and assessing how well they support one another. Additionally, recognising cognitive biases, fostering humility, and practising structured reasoning are crucial for countering flaws in human cognition and navigating an increasingly complex and manipulative information landscape. These skills enable clearer thinking, better decision-making, and meaningful contributions to society.

Implicit premises, or unstated assumptions in arguments, can influence validity, especially when they are controversial or false. To evaluate arguments effectively, implicit premises should be made explicit, allowing for proper assessment of their truth and their connection to the

conclusion. For instance, normative conclusions, which are statements about what should or ought to happen, require at least one normative premise to be valid, as descriptive premises alone cannot justify normative claims. Failing to include these premises leads to logical fallacies, such as deriving "ought" from "is". Reconstructing arguments with all necessary premises, especially for normative conclusions, ensures clarity and validity in reasoning.

Deductive arguments aim to guarantee the truth of their conclusions through the logical structure of their premises. A valid deductive argument ensures that if all premises are true, the conclusion must also be true, regardless of the premises' actual truth. Validity pertains to an argument's structure rather than its content, as seen in examples like "If all cats can fly and Whiskers is a cat, then Whiskers can fly." However, validity alone does not make an argument sound; the premises must also be true. Understanding common valid patterns, such as Modus Ponens, Modus Tollens, disjunctive syllogism, and hypothetical syllogism, alongside recognising formal fallacies like affirming the consequent or denying the antecedent, is essential for effectively evaluating deductive arguments.

Non-deductive arguments, including probabilistic reasoning, enumerative inferences, arguments from samples, analogies, causal reasoning, and inferences to the best explanation, differ from deductive arguments in that they do not guarantee their conclusions but offer varying degrees of likelihood. For example, Sherlock Holmes' reasoning often involves inference to the best explanation, selecting the most plausible cause for observed phenomena while considering alternatives. The strength of non-deductive arguments depends on how well their premises support the conclusion and can vary in degree, unlike the binary validity of deductive arguments. Evaluating such arguments requires assessing factors like sample size, randomness, relevance, and plausibility. While these arguments are inherently open to error, they remain essential for reasoned judgement in addressing uncertainties in everyday reasoning and scientific investigation.

Evaluating arguments involves analysing their structure and the truth of their premises to determine whether they are sound (deductive arguments) or cogent (non-deductive arguments). A sound argument is valid and has true premises, ensuring a true conclusion. Conversely, a cogent argument is strong and has true premises, making its conclusion likely but not guaranteed. Assessing the truth of premises is essential, as it determines whether statements accurately reflect reality. Non-deductive arguments, while cogent, can lead to false conclusions since their premises only make the conclusion probable. Therefore, beyond evaluating cogency, it is crucial to consider external evidence that might challenge the conclusion, highlighting the nuanced relationship between cogency and truth.

CHAPTER 3: REASONING CRITICALLY

Understanding the art of reasoning is essential for navigating the complex world of arguments, debates, and decision-making. In any discussion, whether in academic circles, professional environments, or casual conversations, reasons serve as the foundation upon which claims are built. However, while anything can technically function as a reason, not all reasons are created equal. What makes a reason compelling to one person may not necessarily persuade another, highlighting the subjective nature of persuasiveness.

This chapter aims to unravel the principles that define what constitutes a good reason and differentiate it from a bad one. By focusing on the content and legitimacy of premises rather than just their logical connection to conclusions, this chapter equips readers with tools to critically evaluate arguments and detect reasoning errors. Through an exploration of common fallacies and cognitive biases, readers will gain insight into how arguments are often weakened by insufficient, irrelevant, or ambiguous premises.

A crucial theme of this chapter is the role of context in reasoning. Different arguments operate within distinct frameworks of possibility: logical, physical, and technological. These types of possibilities impose unique constraints and rules on what can be considered valid reasoning. By distinguishing these layers of possibility, we can better assess whether an argument holds up under scrutiny or collapses under its own assumptions.

The chapter also examines the characteristics of valid reasoning: sufficiency, relevance, and clarity. These criteria serve as benchmarks for evaluating premises and identifying fallacies. Moving forward, the chapter delves into the nuanced distinctions between logical consistency, empirical feasibility, and technological innovation, each influencing the persuasiveness of arguments in different ways.

In addition, the chapter explores informal fallacies. These are subtle and often persuasive errors in reasoning that can mislead even the most critical thinker. From hasty generalisations to appeals to emotion, readers will learn to spot these pitfalls and understand why they fail to provide meaningful support for their conclusions.

Lastly, the chapter broadens to include cognitive biases, which are systematic patterns of flawed thinking that shape how we perceive, evaluate, and respond to arguments. Recognising these biases is essential for improving self-awareness, reducing errors in judgement, and fostering more productive discussions.

Through this comprehensive framework, the chapter not only seeks to enhance your analytical skills but also to cultivate intellectual humility, which is the ability to question your assumptions, consider alternative perspectives, and approach reasoning with clarity and fairness. By the end, you will be better equipped to engage in meaningful discourse, construct sound arguments, and evaluate claims with confidence and precision.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand the role of reasons in arguments:** Define reasons and explain their role in supporting claims and conclusions effectively.
- **Evaluate the quality of reasons:** Assess reasons based on the criteria of sufficiency, relevance, and clarity to determine their validity and strength in arguments.
- **Identify reasoning errors:** Recognise common reasoning errors, including fallacies of insufficiency, irrelevance, and ambiguity, and understand why they weaken arguments.
- **Distinguish between types of possibilities:** Explain the differences between logical, physical, and technological possibilities, and analyse how they influence the validity of arguments.
- **Recognise cognitive biases:** Identify key cognitive biases, such as confirmation bias, hindsight bias, and the Dunning-Kruger effect, and understand their impact on reasoning and decision-making.
- **Address cognitive biases:** Apply strategies to mitigate the influence of cognitive biases and improve the objectivity of reasoning.

3.1. TYPES OF POSSIBILITIES

By Marc Chao

In critical thinking and logical reasoning, understanding the types of possibilities helps us evaluate what can or cannot occur within specific contexts. These possibilities are categorised into logical, physical, and technological possibilities, each operating under unique rules and constraints. Together, they form a comprehensive framework for analysing scenarios, testing arguments, and assessing reality.

Logical Possibility

Logical possibility is the most fundamental type, governed by the unchanging laws of logic. A scenario is considered logically possible if it does not entail a contradiction. For instance, the statement "It is raining, and it is not raining at the same time" is logically impossible because it breaks the law of non-contradiction. Logical possibilities focus on the internal consistency of statements and arguments, providing a foundation for deductive reasoning. In deductive reasoning, if the premises are true and the logic is valid, the conclusion must also be true. This concept is critical for ensuring sound arguments, as any logical contradiction undermines the validity of reasoning.

The Origins of the Laws of Logic

The study of logic has its roots in ancient Greece, particularly in the work of Aristotle, who formalised the first comprehensive system of logic. Aristotle introduced what are now known as the three classical laws of thought or the laws of logic: the law of identity, the law of non-contradiction, and the law of excluded middle. These principles serve as the foundation of rational thought and underpin logical possibility.

The Law of Identity

The law of identity asserts that everything is identical to itself. In simpler terms, it means that an object or entity is what it is and cannot be something else. For example, the statement "A cat is a cat" illustrates the law of identity, affirming that the concept of "cat" remains consistent and distinct from other entities. This law ensures that terms and concepts are clear and consistent, making effective communication and reasoning possible. Without it, distinguishing one entity from another would be impossible, leading to confusion and incoherence.

The Law of Non-Contradiction

The law of non-contradiction states that two contradictory propositions cannot both be true at the same time and in the same sense. For example, the statements "It is raining" and "It is not raining" cannot both be true simultaneously. This law is crucial for maintaining logical consistency, as contradictions would invalidate arguments and make rational discourse impossible. It ensures that reasoning stays grounded and free of conflicting conclusions.

The Law of Excluded Middle

The law of excluded middle asserts that for any given proposition, either the proposition is true, or its negation is true. This means that there is no middle ground. For instance, the statement "The light is on" must be either true or false; it cannot be both on and off simultaneously. This principle eliminates ambiguity, ensuring that every proposition has a clear truth value. By doing so, it provides clarity and decisiveness in reasoning, which is essential for making sound judgements.

Evaluating Logical Possibilities

The three laws of logic provide a framework for evaluating the logical possibility of arguments and scenarios. Consider the statement, "A square circle exists". This proposition is logically impossible because it violates the law of non-contradiction, where a shape cannot simultaneously have the defining properties of both a square and a circle. By applying these laws, we can assess whether scenarios and arguments are logically consistent.

Logical possibility is the cornerstone of rational thinking. It enables us to identify and discard impossible propositions while constructing coherent and valid arguments. This foundation ensures that our reasoning is robust, reliable, and free from contradictions, empowering us to make clear, well-supported decisions.

Physical Possibility

Physical possibility refers to what can happen within the constraints of the natural laws that govern the universe. These laws, derived from centuries of scientific observation and experimentation, explain how matter and energy interact and behave. A scenario is physically possible if it does not violate these well-established principles. For example, while it is easy to imagine someone flying unaided (a logical possibility), it is not physically possible under the known laws of gravity and human physiology.

The Laws of Nature

The laws of nature serve as the foundation for understanding physical possibilities. These laws are universal and unchanging, providing a reliable framework for analysing how the physical world works. Examples include:

- **Newton's Laws of Motion:** These laws describe how objects move and interact with forces. For instance, Newton's first law (the law of inertia) states that an object will remain at rest or in uniform motion unless acted upon by an external force.
- **Laws of Thermodynamics:** These govern energy transformations. The first law states that energy cannot be created or destroyed, only converted from one form to another, while the second law explains that entropy in a closed system always increases over time, ruling out perpetual motion machines.
- **Quantum Mechanics:** This field addresses the behaviour of particles at atomic and subatomic levels. Principles like wave-particle duality and the uncertainty principle are central to this branch of physics.
- **Theory of Relativity:** Einstein's theories describe how space, time, and gravity are interconnected. Special relativity addresses the speed of light as a universal constant, while general relativity explains gravitational effects on spacetime.

These laws help scientists predict phenomena and determine the boundaries of physical possibility.

Evaluating Physical Possibilities

Understanding physical possibility allows us to distinguish between what is achievable in the natural world and what exists only in imagination. For example, the idea of a perpetual motion machine is logically conceivable but physically impossible due to the laws of thermodynamics. Similarly, while faster-than-light travel seems exciting in fiction, it violates Einstein's theories, making it physically impossible with our current understanding.

Physical possibilities are essential for scientific progress. They provide the basis for formulating testable hypotheses and conducting experiments. Scientists rely on these laws to validate their ideas, ensuring they align with empirical evidence. For instance:

- **Space travel:** The development of rockets and spacecraft relies on principles such as Newton's third law (for every action, there is an equal and opposite reaction).
- **Medical technology:** Innovations like MRI machines and laser surgery operate within the laws of electromagnetism and optics.
- **Renewable energy:** Solar panels work based on the principles of photovoltaics, converting light into electrical energy.

These examples demonstrate how physical possibilities translate into groundbreaking advancements that align with the laws of nature.

The Role of Physical Possibility in Scientific Progress

Recognising physical possibilities helps scientists and innovators understand the limitations imposed by natural laws and identify realistic goals. While faster-than-light travel is currently deemed impossible, theoretical explorations of concepts like wormholes and warp drives continue to challenge these limits, showing how physical constraints can inspire creativity.

By focusing on what is physically possible, scientists can channel their efforts into achievable objectives, leading to new technologies and a deeper understanding of the universe. This disciplined approach ensures that scientific theories are grounded in empirical reality and not mere speculation.

Understanding physical possibility is a cornerstone of critical thinking. It allows us to evaluate the feasibility of scenarios and hypotheses within the bounds of natural laws. This helps us separate plausible ideas from those that are purely imaginative. Additionally, it fosters innovative problem-solving by encouraging thinkers to work creatively within realistic constraints, promoting evidence-based reasoning and practical decision-making.

Technological Possibility

Technological possibility refers to what can be achieved through the development and application of technology, whether current or in the future. Unlike logical or physical possibilities, technological possibilities are dynamic and constantly evolving as new scientific discoveries and engineering breakthroughs occur. For example, travelling to the moon was once a dream relegated to science fiction, but with advancements in space exploration technology, it became a reality in 1969 with the Apollo 11 mission. Technological possibilities are essential for driving innovation, pushing the boundaries of what humanity can achieve, and opening up new opportunities for progress in fields like engineering, computer science, and medicine. They often serve as the catalysts for breakthroughs that significantly improve the quality of life.

The Dynamic Nature of Technological Possibility

One of the defining features of technological possibility is its ever-changing nature. What seems impossible today can become feasible tomorrow through human ingenuity and scientific advancements. For instance, the concept of instant communication across vast distances was once inconceivable, but with the invention of the telephone, and later the internet, it has become a fundamental part of modern life. Technological possibilities expand with each new discovery, redefining what is achievable and inspiring further innovation.

Historical milestones illustrate this dynamic nature vividly. Consider space travel: in the early 20th century, the idea of venturing beyond Earth's atmosphere seemed outlandish. However, the launch of Sputnik in 1957 marked the dawn of the space age, and the moon landing in 1969 cemented space exploration as a technological reality. Similarly, the evolution of computing power showcases the rapid advancement of technology. Early computers were room-sized machines with limited capabilities, but today, smartphones fit in our pockets and boast computational power that far exceeds that of the computers used in the Apollo missions. In medicine, innovations like MRI machines have transformed diagnostics, enabling non-invasive imaging, while breakthroughs in biotechnology, such as CRISPR gene-editing technology, hold promise for curing genetic disorders.

Evaluating Technological Possibilities

Understanding technological possibilities is crucial for critical thinking and decision-making. It allows us to assess the feasibility of innovations and predict their potential impact on society. This forward-thinking approach encourages us to consider not only what is achievable now but also what might become possible with future advancements. For instance, while faster-than-light travel remains physically impossible under current scientific understanding, technological exploration of concepts like wormholes or warp drives reflects how technological possibilities push the boundaries of our imagination and understanding.

Evaluating technological possibilities also requires considering the interplay with logical and physical possibilities. For example, teleportation is logically conceivable because it involves no inherent contradictions. Physically, however, it may conflict with natural laws, such as the conservation of energy and matter. Even if it were physically possible, the technological feasibility of teleportation would depend on advancements in energy management, materials, and cost-effective implementation, none of which currently exist. Similarly, the development of artificial intelligence (AI) is logically and physically possible, but technological challenges such as computational power, ethical concerns, and practical applications remain hurdles to its full realisation.

The Framework of Possibilities

By distinguishing between logical, physical, and technological possibilities, we can approach complex problems from multiple perspectives. Logical impossibilities cannot translate into physical or technological realities, while physical possibilities must also adhere to logical consistency. Technological possibilities, on the other hand, require alignment with both logical and physical realities, and their feasibility often hinges on factors like cost, resources, and practicality. This layered framework helps ensure that our arguments, hypotheses, and decisions are robust, grounded in reality, and achievable within known constraints. By applying this framework, we can navigate the complexities of modern challenges and innovate responsibly, advancing our understanding and capabilities in meaningful ways.

3.2. INFORMAL FALLACIES

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

A fallacy is a flaw in reasoning that undermines the validity of an argument. This flaw can arise from either the structure of the argument (how its propositions are arranged and connected) or its content (what the propositions actually state). While the form of an argument always matters, informal fallacies specifically deal with errors in how the content of propositions is handled as premises or reasons. This differs from formal fallacies, which we discussed in the previous chapter, and which concern the structural relationships between statements and conclusions.

Informal fallacies typically occur when a premise fails to adequately support a conclusion because it is insufficient, irrelevant, or ambiguous. Unlike formal fallacies, informal fallacies cannot be identified simply by examining the argument's structure; they require an understanding of the content and context of the premises.

It might surprise you to learn just how common these fallacies are, both in casual conversations and formal debates. For a variety of psychological reasons, fallacious reasoning often seems persuasive, even when it should not be. Part of what makes fallacies so pervasive is their ability to appear convincing on the surface, even though they lack real substance. However, after working through the concepts in previous chapters, you should now be better equipped to recognise and avoid these misleading patterns of reasoning.

Why Study Informal Fallacies?

The list of informal fallacies we will explore in this chapter is not exhaustive. In fact, no single list could ever capture every possible fallacious reason; there are hundreds of known examples, and new ones emerge as our ways of reasoning and communicating evolve. The purpose of this list is not to have you memorise obscure Latin names or classify every bad argument you encounter. Instead, our goal is to help you understand the principles behind why these fallacies fail as reasons so you can recognise them when they appear, regardless of their label.

Organising Informal Fallacies

There is no universally agreed-upon way to classify informal fallacies. Different systems categorise them in various ways. Some may use four categories, while others may organise them differently. Additionally, certain fallacies might comfortably fit into more than one category. For example, a fallacy categorised as insufficient reasoning might also exhibit elements of irrelevance.

Despite these overlaps, organising fallacies into groups can still be helpful because it highlights the general principles that underlie faulty reasoning. For simplicity and clarity, we will focus on a three-category system:

1. Insufficient Reasons – Premises that fail to provide enough support for the conclusion.
2. Irrelevant Reasons – Premises that may seem related but do not actually support the conclusion.
3. Ambiguous Reasons – Premises that are unclear, vague, or open to multiple interpretations.

This three-group approach provides a clear and practical framework for identifying and analysing informal fallacies. Rather than memorising labels, focus on understanding why these types of reasoning fail and how they can be avoided. With these tools, you will be better prepared to assess arguments critically, spot weak reasoning, and construct stronger, more persuasive arguments of your own.

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

3.3. FALLACIES OF INSUFFICIENCY

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Fallacies of insufficiency occur when the evidence or reasons provided in an argument are inadequate to support the conclusion. In such cases, the premises fail to deliver the necessary support or justification, leaving the conclusion resting on weak or unjustified assumptions. In any argument, the burden of proof lies with the person presenting the claim, requiring them to provide sufficient evidence and sound reasoning to make their conclusion credible. When this standard is not met, the argument falls into the trap of insufficient reasoning.

Many fallacies of insufficiency stem from poorly constructed inductive arguments, including predictive inductions, generalisations, analogies, and cause-and-effect reasoning. These forms of reasoning, when properly executed, can be powerful tools for drawing conclusions from evidence. However, if the premises lack sufficient detail, breadth, or accuracy, the resulting argument will be weak and unreliable.

What makes these fallacies particularly notable is that they often present their premises as though they should be convincing in isolation, without acknowledging the gaps in reasoning that weaken the argument. In many cases, these arguments are not entirely false, they are simply incomplete. With additional premises, more robust evidence, or clearer reasoning, many of these arguments could potentially be strengthened and made persuasive.

Hasty Generalisation

The hasty generalisation fallacy occurs when someone draws a broad conclusion about an entire group or category based on an inadequate, biased, or unrepresentative sample. Inductive reasoning often involves making generalisations from specific examples, but if the sample is too small, poorly selected, or otherwise flawed, the resulting conclusion lacks credibility and cannot be logically defended.

For instance, if someone visits one restaurant in a new city, has a bad experience, and then declares, “All restaurants in this city are terrible”, they are committing a hasty generalisation. The sample size in this case, a single restaurant, is far too small to justify such a sweeping conclusion.

This fallacy is particularly problematic because it often serves as the foundation for stereotypes and prejudice. Stereotypes frequently emerge from our tendency to cling to weak evidence or overextend limited observations to represent entire groups. This behaviour is often motivated by a desire to reduce uncertainty or simplify complex realities. When confronted with limited information, the human mind is prone to jumping to conclusions rather than seeking more robust evidence.

Hasty generalisations are widespread in both casual conversations and formal arguments. The fallacy also manifests in several common variations, including Insufficient Sample, where conclusions are drawn from too few examples; Converse Accident, where a general rule is misapplied to an exceptional case; Faulty Generalisation, where broad claims are made without sufficient evidence; Biased Generalisation, where the sample is unrepresentative of the larger group; and Jumping to Conclusions, where premature assumptions are made without adequate evidence. Despite their differences in appearance, these variations all share the same fundamental flaw: the evidence provided is simply insufficient to support the conclusion.

Recognising the hasty generalisation fallacy is essential for developing stronger reasoning skills. Effective inductive reasoning requires attention to sample size, representativeness, and the reliability of the evidence being presented. Before drawing broad conclusions, it is crucial to evaluate whether the examples used are sufficient and representative of the larger context.

By remaining mindful of these factors, we can avoid making sweeping claims based on weak premises. Instead, our arguments will be grounded in logic, well-supported by evidence, and ultimately far more persuasive and credible. Understanding and avoiding hasty generalisations not only strengthens our reasoning but also helps prevent the spread of harmful stereotypes and misinformation.

Figure 3.3.1 shows a few more examples of hasty generalisation:

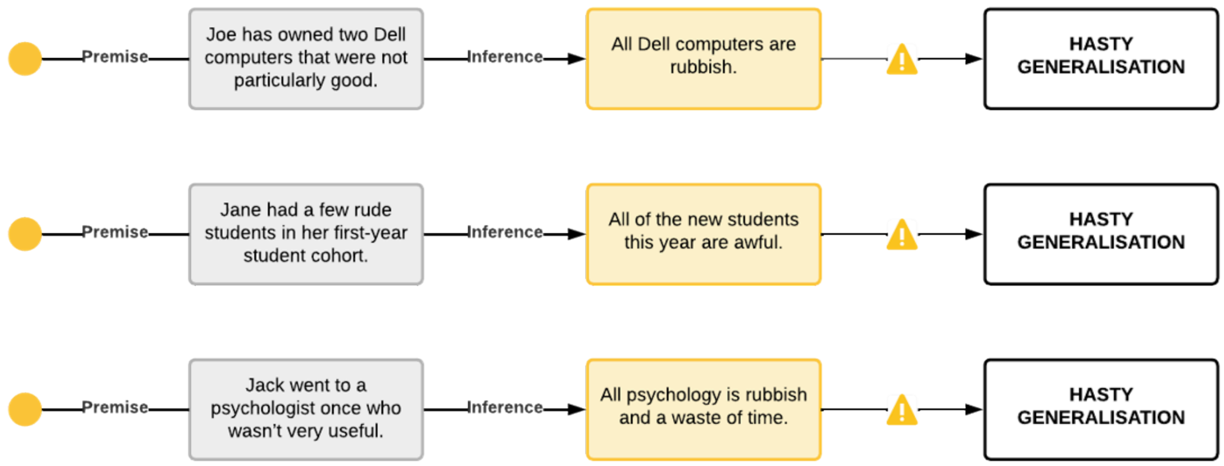


Figure 3.3.1. Additional examples of hasty generalisation by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Post Hoc Ergo Propter Hoc (False Cause)

The Post Hoc Ergo Propter Hoc fallacy, often shortened to post hoc fallacy, arises from the mistaken belief that because one event follows another, the first event must have caused the second. The Latin phrase translates to "after this, therefore because of this", capturing the core assumption of this reasoning error. While identifying cause-and-effect relationships is a central goal of inductive reasoning, establishing such relationships with certainty is notoriously difficult. The post hoc fallacy represents a common pitfall in causal reasoning, where causation is inferred without sufficient evidence to justify the conclusion.

For an argument to credibly assert a cause-and-effect relationship, it must satisfy three key conditions. First, there must be a correlation, which is a measurable relationship between the two phenomena. Second, there must be temporal order, meaning the supposed cause must occur before the effect because an effect cannot precede its cause. Third, alternative causes or confounding factors must be ruled out, ensuring that other possible explanations are not misleadingly attributed to the relationship.

In strong inductive arguments about causality, the premises must address all three of these conditions. When one or more of these criteria are ignored or insufficiently supported, the reasoning becomes fallacious, and the argument falls into the trap of post hoc reasoning.

At its root, this fallacy represents a premature assumption of causality. It occurs when someone jumps to the conclusion that one event caused another simply because they happened in sequence. However, correlation does not equal causation, and two events occurring consecutively may have no meaningful connection whatsoever.

This fallacy is frequently seen in superstitious thinking or casual observations where patterns are misinterpreted as causal links. For example, someone might say, "I wore my lucky socks, and my team won the game. Therefore, my socks caused the victory." Or another might claim, "Every time I wash my car, it rains the next day. Washing my car must cause rain." In both cases, the reasoning is flawed because the arguments fail to eliminate alternative explanations or demonstrate a genuine causal connection between the events.

Recognising and understanding the post hoc fallacy is essential for evaluating causal arguments critically. Just because two events occur in succession does not mean one caused the other. Proper reasoning requires us to look beyond simple correlations and carefully examine whether the relationship satisfies the three key conditions for causality. Without this scrutiny, we risk drawing misleading conclusions based on coincidence or superficial patterns rather than on solid evidence and sound reasoning.

Figure 3.3.2 shows a few more examples of the post hoc fallacy:

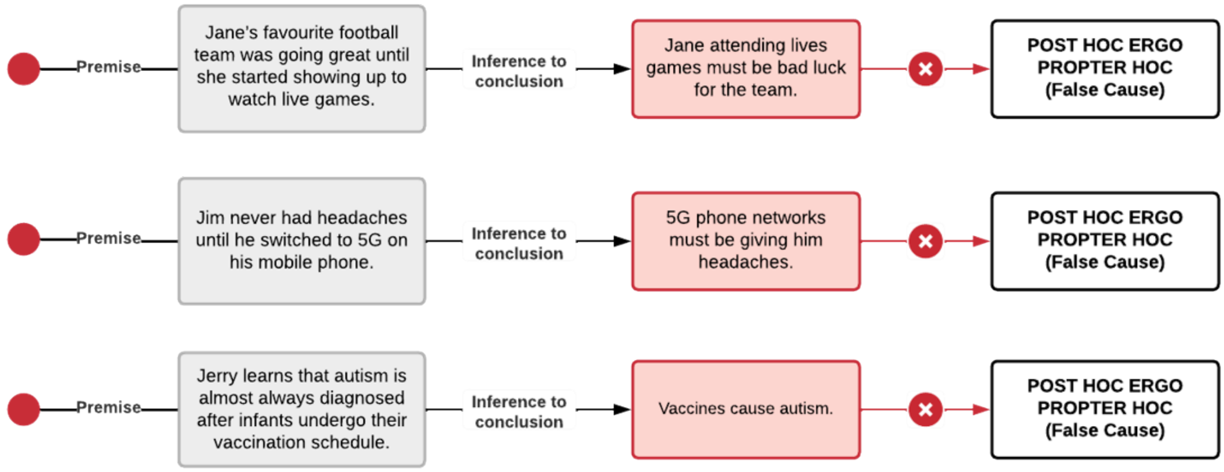


Figure 3.3.2. Post Hoc Ergo Propter Hoc Examples by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Slippery Slope

The slippery slope fallacy occurs when an argument predicts a chain of future events without sufficient evidence to support the claim. In these scenarios, one initial action is presented as the trigger for an unstoppable series of increasingly severe or extreme consequences. However, predictions about future outcomes must be backed by their own reasoning and evidence, and they cannot simply be assumed or asserted without justification.

These arguments often rely on a “give an inch, take a mile” style of reasoning, appealing to the imagery of a domino effect or chain reaction where one event inevitably leads to another, often disastrous, outcome. In many cases, slippery slope arguments are accompanied by appeals to fear or appeals to inevitability, focusing on worst-case scenarios rather than presenting a logical, evidence-based progression from one event to the next.

It is important to recognise that not all slippery slope arguments are fallacious. There are situations where a sequence of events is genuinely plausible and well-supported by evidence. For example, in a hypothetical syllogism, a chain of reasoning can be valid if each step logically follows from the previous one and is supported by clear evidence. The fallacy arises specifically when the argument overreaches, making assumptions about future outcomes without adequate reasoning or supporting evidence.

For instance, someone might claim, “If we allow one student to hand in their assignment late, soon everyone will start missing deadlines, and eventually, academic standards will collapse entirely.” At first glance, this argument seems plausible, but upon closer examination, it becomes clear that no evidence is provided to support the claim that leniency in one instance will trigger widespread academic decline. However, if the speaker were to present data showing past examples where such leniency led to systemic issues with deadlines and accountability, the argument would transition from being fallacious to plausible, becoming a legitimate concern supported by evidence.

The key difference lies in whether the chain of events is supported by logical reasoning and evidence or merely assumed through speculation and fear-mongering. Valid slippery slope arguments are carefully constructed, showing clear causal links between each step, while fallacious ones rely on exaggeration and emotional appeal rather than well-reasoned analysis.

Understanding this distinction is essential for evaluating slippery slope claims effectively. While it is wise to consider potential consequences of actions, it is equally important to demand evidence for each step in the predicted sequence. Without this evidence, slippery slope arguments remain speculative and unconvincing, serving more as rhetorical devices than as reliable reasoning. Recognising when this fallacy is at play helps ensure that discussions remain focused on evidence and logic rather than being derailed by unfounded assumptions or exaggerated fears.

Figure 3.3.3 shows a few more examples of the slippery slope fallacy:

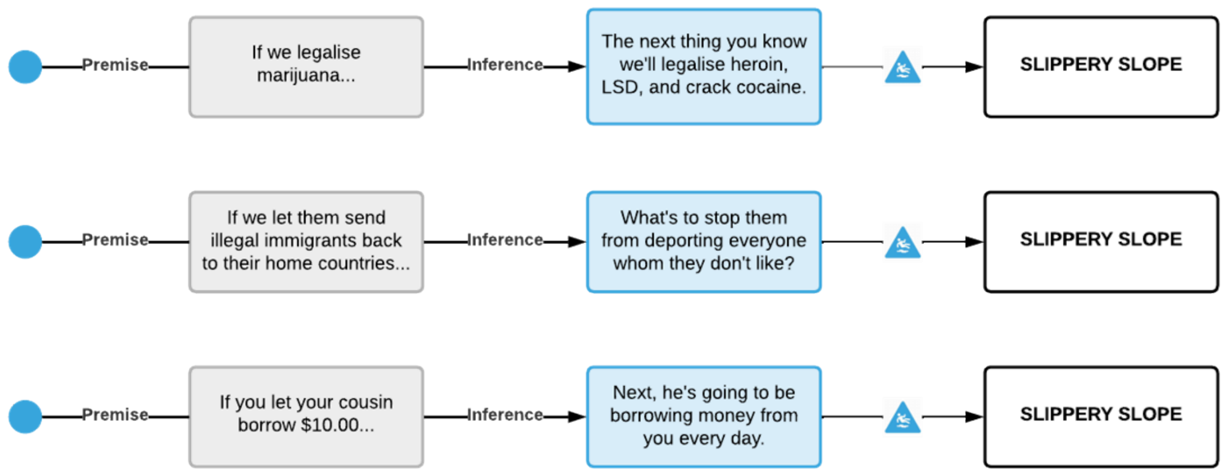


Figure 3.3.3. Additional examples of the slippery slope fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Faulty or Weak Analogy

A faulty or weak analogy occurs when an argument compares two things that are not sufficiently similar in relevant ways to support its conclusion. Like causal fallacies and hasty generalisations, this error represents a breakdown in inductive reasoning, specifically in the use of analogies to infer conclusions. While analogies can be a powerful tool in reasoning, they become fallacious when the similarities between the compared items are superficial, irrelevant, or insufficient to justify the argument's conclusion.

In inductive reasoning, an analogy operates by suggesting that because two things are alike in certain respects, they must also be alike in other respects. However, if the shared similarities are trivial or unrelated to the core point being argued, the analogy fails to provide valid support for the conclusion.

For example, someone might say: "Employees are like nails. Just as nails must be hit on the head to work properly, employees must be managed with strict discipline to be effective." At first glance, this comparison might appear clever or even insightful. However, the similarity between nails and employees is superficial and irrelevant to the argument about workplace management. Upon closer inspection, it becomes clear that employees are complex human beings, capable of thought, creativity, and emotion, while nails are inanimate objects with no agency or capacity for reasoning. The analogy collapses under scrutiny because the comparison is based on irrelevant similarities and ignores crucial differences.

This fallacy is known by several other names, including Bad Analogy, False Analogy, Questionable Analogy, Argument from Spurious Similarity, and False Metaphor. Despite the variety of terms, they all describe the same fundamental error: drawing conclusions from a comparison where the similarities are either superficial or unrelated to the argument's purpose.

While analogies can serve as effective tools for clarifying ideas, illustrating concepts, and supporting arguments, they must always be evaluated critically. The strength of an analogy hinges on whether the similarities are significant and relevant to the conclusion being drawn. If they are not, then no matter how persuasive or clever the comparison might initially seem, the argument remains fundamentally flawed.

In essence, strong analogies rely on meaningful and relevant similarities, while weak analogies fail because their comparisons are superficial or unrelated to the argument's central point. Recognising this distinction is key to using analogies effectively and avoiding fallacious reasoning. By approaching analogies with a critical eye, we can ensure they serve as reliable tools for reasoning rather than misleading rhetorical devices.

Figure 3.3.4 shows a few more examples of the weak or faulty analogies:

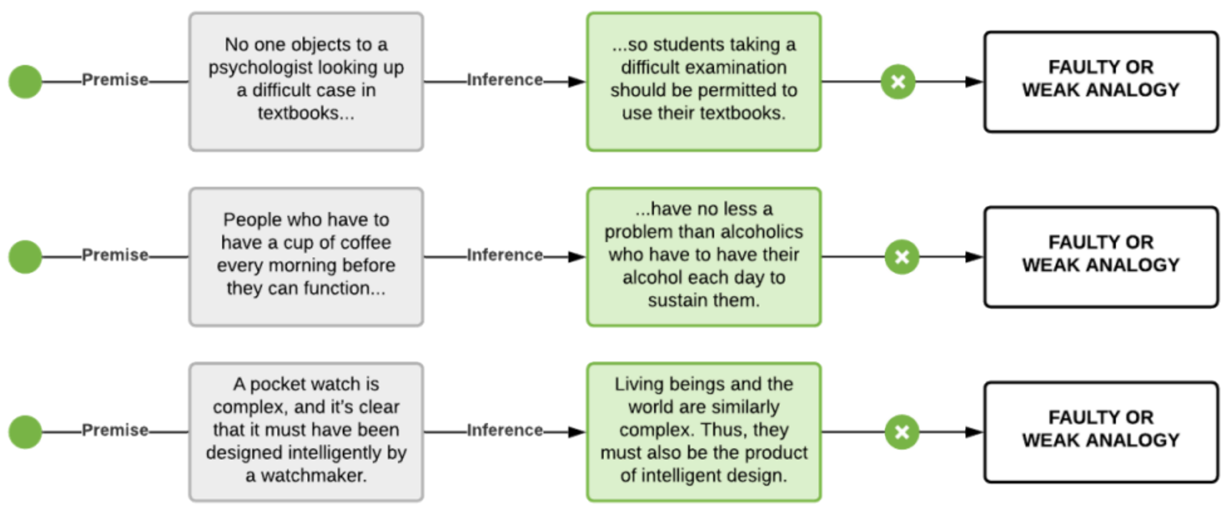


Figure 3.3.4. Additional examples of the faulty or weak analogy fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Argumentum ad Verecundiam (Appeal to Authority, Unqualified)

The appeal to authority fallacy, also known as argumentum ad verecundiam, occurs when someone cites an authority figure to support a claim, but the authority in question is either unqualified or irrelevant to the topic at hand. While this fallacy could also be classified under fallacies of irrelevance, it often reflects insufficient reasoning because it relies on the credibility of an authority figure without offering supporting evidence or a valid rationale for the claim.

At its core, this fallacy assumes that if an authority figure believes something, it must be true. However, history offers numerous examples of brilliant individuals holding incorrect or questionable beliefs. For instance, Isaac Newton, one of the greatest scientific minds in history, dedicated significant time to alchemy and apocalyptic predictions, pursuits that lacked scientific validity. Similarly, while Albert Einstein's political opinions might be thought-provoking, they are not inherently more credible than those of a political scientist. Quoting Einstein to support a political argument would be as misguided as quoting a political theorist on the nuances of special relativity theory.

This fallacy is especially common in social media debates, where individuals frequently invoke famous names to bolster their arguments, regardless of whether those figures have relevant expertise in the subject being discussed. The assumption seems to be that fame or success in one domain automatically translates to authority in all others, which is rarely the case.

It is essential to clarify that not all appeals to authority are fallacious. When done correctly, referencing an authority figure can serve as a shorthand for appealing to the current state of knowledge, scientific consensus, or credible expertise. However, the strength of such an argument does not stem from the individual making the claim, but from the evidence and reasoning they represent.

For instance, citing Stephen Hawking's views on black holes is not fallacious because his perspective represents decades of rigorous scientific research and consensus. Similarly, trusting medical advice from a qualified doctor is not an appeal to blind faith but rather reliance on their training, expertise, and evidence-based knowledge. In both examples, the authority figure serves as a conduit for established evidence and knowledge, rather than being the sole justification for the claim.

The appeal to authority fallacy typically emerges in two scenarios. The first is irrelevant expertise, where the authority figure cited lacks expertise in the specific subject being discussed. For example, relying on a physicist's opinion on nutrition science would be inappropriate because their area of expertise does not extend to dietary research. The second scenario involves a lack of verification, where the authority figure's expertise is either not properly substantiated or no additional evidence is provided to justify why their opinion should be trusted.

For example, someone might say, "97% of climate scientists believe in human-caused climate change, so it must be true." While this claim references expert consensus and is not technically fallacious, it is unpersuasive on its own. A stronger argument would explain why these scientists believe this, referring to the evidence and reasoning supporting their consensus rather than relying solely on the percentage figure.

The key takeaway is that appeals to authority are not inherently fallacious, but they become weak arguments when they rely solely on the reputation of the authority figure rather than the evidence they represent. Whenever possible, it is better to focus on the reasoning and evidence behind a claim rather than the authority delivering it. While authority can add credibility to an argument, it should never replace clear reasoning and verifiable evidence. Recognising this fallacy helps ensure that discussions remain grounded in sound reasoning rather than misplaced reliance on perceived authority.

Figure 3.3.5 shows a few more examples of the appeal to authority fallacy:

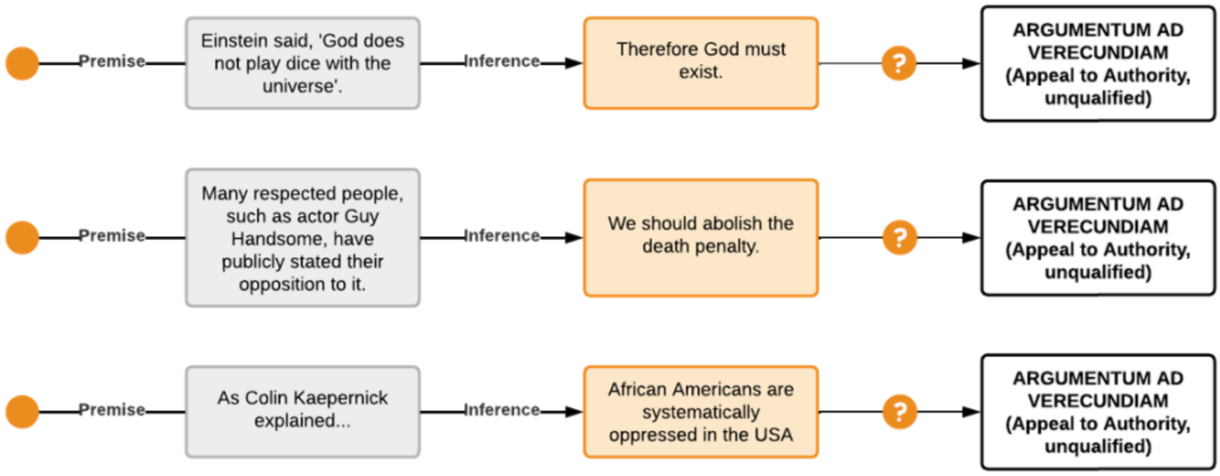


Figure 3.3.5. Additional examples of the appeal to authority fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Argumentum ad Ignorantiam (Appeal to Ignorance)

The appeal to ignorance (argumentum ad ignorantiam) occurs when someone argues that a claim must be true simply because there is no evidence proving it false, or conversely, false because there is no evidence proving it true. Rational thinking does not operate on the assumption that something is true until proven false or false until proven true. A lack of evidence is not evidence in itself and cannot be used to justify any positive conclusion.

This fallacy is often intertwined with the shifting of the burden of proof, another common reasoning error. In any argument, the responsibility to provide evidence lies with the person making the positive claim. Simply stating, "No one has disproven my claim", does not qualify as valid support for that claim. The obligation to provide reasoning and evidence cannot be transferred to someone else.

You will frequently encounter this fallacy in pseudoscience and certain commercial industries, such as dieting, supplements, and beauty products. These fields often rely on the absence of disproof rather than presenting robust evidence to substantiate their claims. For example, a beauty product might claim to reduce wrinkles and add, "No study has proven otherwise". Such reasoning shifts the focus away from the lack of supporting evidence and places an unfair expectation on others to disprove the claim.

This fallacy also tends to arise when people hold strong personal beliefs or have an emotional attachment to an idea. When someone feels deeply invested in a claim, they may demand that others disprove it instead of offering their own evidence to support it. For instance, an astrology enthusiast might argue, "You can't prove astrology doesn't work, so it must be true." This reasoning is flawed because the responsibility to provide evidence always rests with the person making the positive claim. In such situations, the most rational response is to remain sceptical and withhold belief until sufficient evidence is provided.

However, it is important to note that in certain contexts, a lack of evidence can be meaningful, particularly in scientific research. When scientists actively test a hypothesis and repeatedly fail to find supporting evidence despite rigorous attempts, this absence of evidence becomes significant and can be a valid reason for rejecting the hypothesis.

For example, consider the claim: "There's no evidence that childhood vaccinations are linked to autism." This statement is not an appeal to ignorance because scientists have spent decades conducting rigorous studies on this hypothesis. Despite extensive research, no credible evidence has been found to support the claim. In this case, the absence of evidence is not due to a lack of investigation, but rather a consistent pattern of negative results. Accepting this conclusion is therefore rational and justified.

The appeal to ignorance fallacy happens when someone uses a lack of evidence as proof of their claim, rather than presenting positive evidence to support it. Rational reasoning requires that those making a claim bear the burden of proof and provide clear, verifiable evidence for their position. While scientific findings sometimes rely on an absence of evidence, this approach is only valid when a thorough investigation has been conducted and no supporting data has been found despite consistent effort.

A lack of evidence is not the same as evidence of absence, unless it is backed by a thorough investigation and consistent findings. In situations where evidence is unclear or incomplete, scepticism and critical thinking remain essential tools for evaluating such claims responsibly.

Figure 3.3.6 shows a few more examples of the appeal to ignorance fallacy:

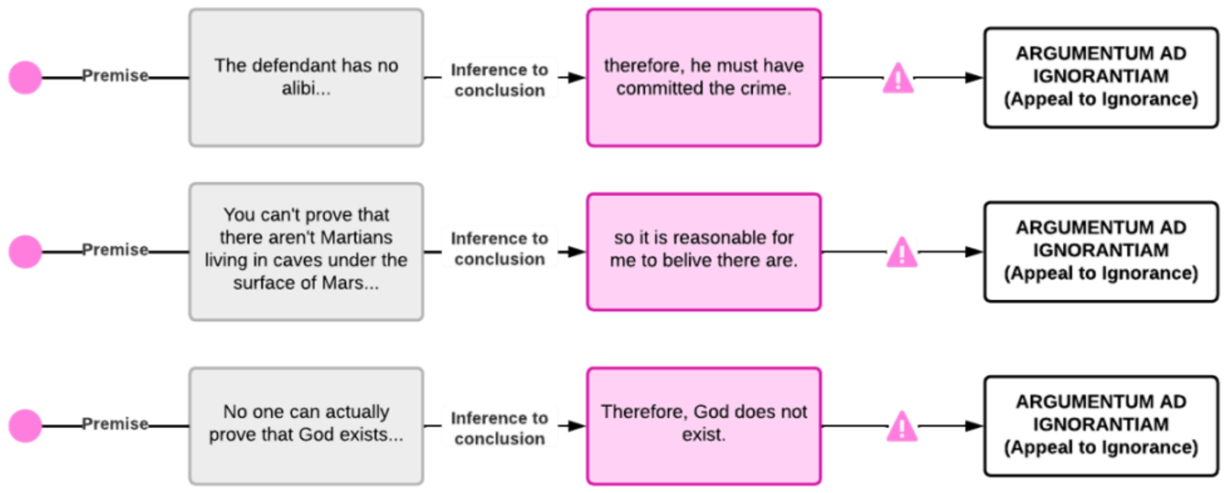


Figure 3.3.6. Additional examples of the appeal to ignorance fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Petitio Principii (Begging the Question)

The Petitio Principii fallacy, commonly known as begging the question, occurs when an argument's premises assume the truth of its conclusion rather than offering independent evidence to support it. In essence, the conclusion is subtly embedded or rephrased within the premises, creating a circular reasoning pattern that offers an illusion of support without introducing any new information.

This fallacy is often misunderstood in casual conversation. In everyday language, people use "begs the question" to mean "raises the question", as in: "John is smart, and it begs the question: Why is he with that girl?" However, in the context of logical reasoning, begging the question specifically refers to an argument where the premise and conclusion are essentially saying the same thing in different words.

In circular reasoning, the argument creates a loop where the premise depends on the conclusion being true and vice versa. While this structure might sound persuasive, it ultimately fails to provide meaningful support for the claim being made. For example, consider the claim: "The new health supplement is effective because it's the most popular product on the market for improving energy levels." Here, the premise (it's the most popular product on the market for improving energy levels) assumes the conclusion (the health supplement is effective) is already true. At the same time, the conclusion (the health supplement is effective) is justified by the premise (its popularity). This creates a circular relationship where no independent evidence supports the supplement's effectiveness. Another example is the statement: "Everyone wants the new iPhone because it's the hottest new gadget on the market." In this case, the premise (it's the hottest new gadget on the market) is essentially a reworded version of the conclusion (everyone wants it). No external reasoning or evidence is provided to explain why the iPhone is desirable, resulting in circular reasoning. In both examples, the premises merely restate the conclusion in slightly altered language, failing to offer any genuine support or meaningful justification.

This fallacy is also referred to by other names, including Circular Argument, Circulus in Probando, and Vicious Circle. Each of these terms points to the same flaw: using the conclusion as evidence for itself rather than providing independent support.

Circular arguments are fallacious because they fail to advance reasoning or introduce new evidence. They may appear persuasive on the surface because of their repetitive structure, but ultimately, they lack the foundation needed for a sound argument. When evaluating an argument for this fallacy, consider the following: Is the premise offering independent support for the conclusion, or is it just rephrasing it? And, if the premise were removed, would the conclusion still hold up on its own?

To address circular reasoning, it is helpful to ask for independent evidence that does not rely on the conclusion being true. Clarifying the structure of the argument can also reveal whether the premise and conclusion are genuinely distinct claims. Additionally, focusing on external justification in the form of facts, data, or logic that exist outside of the premise–conclusion loop helps prevent this fallacy from undermining meaningful discussion.

The Petitio Principii fallacy ultimately undermines logical reasoning by recycling the conclusion as a premise, creating an endless loop of unsupported reasoning. Effective critical thinking requires the ability to identify circular reasoning patterns, demand independent evidence, and ensure that premises provide genuine support for conclusions rather than simply rephrasing them.

Figure 3.3.7 shows a few more examples of the begging the question fallacy:

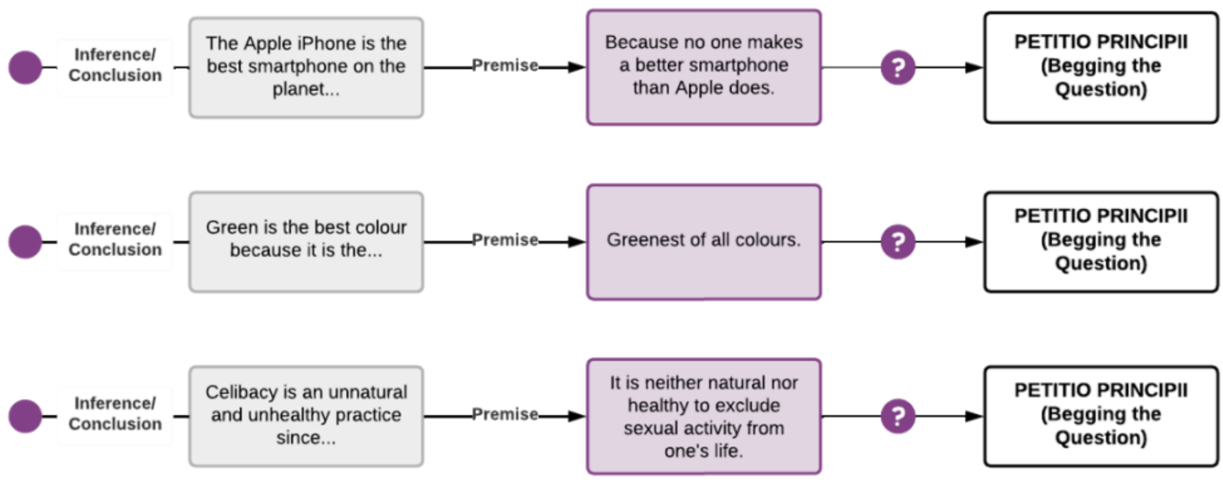


Figure 3.3.7. Additional examples of the begging the question fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

False Dichotomy

A false dichotomy occurs when an argument presents only two mutually exclusive options as if they are the only possible choices, even though there are usually many more alternatives available. This fallacy oversimplifies complex issues, reducing them to a black-and-white scenario while ignoring the nuance and variety of real-world possibilities.

In many instances, false dichotomies are a deliberate rhetorical strategy used to manipulate or limit an audience's perception of their available choices. By framing a situation as an "either/or" decision, the speaker attempts to make their preferred choice seem more reasonable or compelling while dismissing or hiding other potential alternatives.

This fallacy often reveals itself through key phrases like "the only alternative" or through the frequent use of the word "or" in the argument's premises. It is a particularly common tactic in political speeches, where speakers simplify multifaceted policy issues into two extreme and seemingly opposed choices to sway public opinion or galvanise support.

For example, someone might say, "You're either with us, or you're against us", or claim, "We must either increase surveillance or face total chaos in society". In both cases, the argument artificially limits the options to only two extremes, ignoring the possibility of more balanced or alternative approaches that might exist between or beyond the presented choices.

The false dichotomy fallacy is known by several other names, including False Dilemma, All-or-Nothing Fallacy, Either/Or Fallacy, Black-and-White Thinking, Polarisation, Fallacy of False Choice, Fallacy of Exhaustive Hypotheses, No Middle Ground, and Bifurcation. Despite these different names, they all refer to the same fundamental error: reducing a complex situation to an oversimplified binary choice.

False dichotomies are problematic because they limit critical thinking and oversimplify complex problems. They force people into making decisions based on artificially restricted choices, preventing them from considering alternative solutions or exploring middle-ground positions that might offer better outcomes.

To identify and counter a false dichotomy, it is important to ask whether there are other possibilities beyond the two presented options. Consider whether the choices are genuinely mutually exclusive or if they might coexist or overlap in some way. It is also useful to examine whether the argument relies on an oversimplification of an inherently complex issue.

Recognising and challenging false dichotomies helps to prevent being cornered into false choices and allows for more thoughtful and nuanced reasoning. By doing so, we can approach complex issues with clarity and openness, avoiding the pitfalls of oversimplification and exploring the full range of available possibilities before making decisions.

Figure 3.3.8 shows a few more examples of the false dichotomy fallacy:

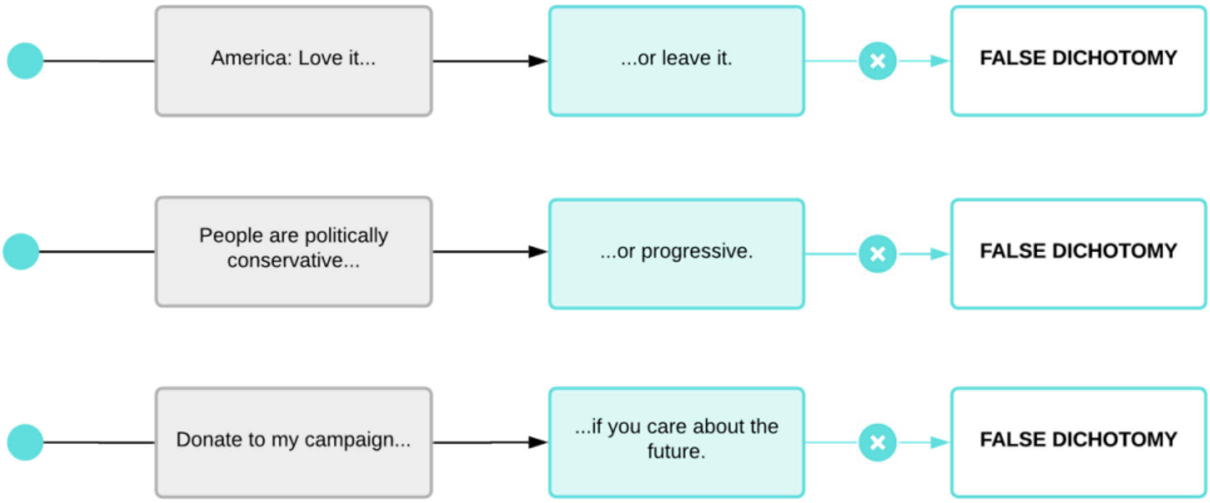


Figure 3.3.8. Additional examples of the false dichotomy fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

---

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

3.4. FALLACIES OF IRRELEVANCE

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Fallacies of irrelevance differ from fallacies of insufficiency in a fundamental way. While insufficient reasoning arises when an argument lacks key evidence or supporting details, the premises at least attempt to address the claim directly. In contrast, irrelevant reasons introduce premises that have no meaningful connection to the claim being made, effectively shifting the focus away from the central argument rather than supporting it.

The core issue with fallacies of irrelevance is that they distract attention from the original claim by introducing unrelated information. This distraction can derail meaningful discussion and prevent proper evaluation of the argument. Determining whether a premise is relevant or irrelevant depends entirely on the specific claim under discussion, meaning relevance must always be evaluated on a case-by-case basis.

For instance, it would be fallacious to dismiss vegetarianism simply because an infamous historical figure like Hitler was a vegetarian. In this example, the premise (Hitler's dietary choices) has no bearing on the ethical or health arguments for vegetarianism. The historical association serves as an emotional distraction rather than addressing the merits of the claim itself.

However, it would not be fallacious to criticise Hitler's human rights policies based on his history of cruelty and moral failure. In this context, the premise (Hitler's actions) is directly relevant to evaluating his policies because it provides contextual and moral insight into their implications.

One specific example of this type of fallacy is so frequent and recognisable that it has been given a name: "Reductio ad Hitlerum". Sometimes humorously referred to as "playing the Nazi card", this fallacy occurs when someone dismisses or discredits an idea solely because an infamous or disliked figure once held a similar view. Instead of engaging with the actual merits of the argument, the discussion is diverted to an irrelevant association, rendering the reasoning unproductive and logically flawed.

Argumentum ad Hominem (Against the Person)

The ad hominem fallacy occurs when someone attempts to dismiss or undermine an argument by attacking the person making the argument rather than addressing the argument itself. The term ad hominem translates from Latin as "against the man", emphasising its focus on the individual's character, motives, or personal traits instead of engaging with the reasoning or evidence they present.

This fallacy arises when someone tries to discredit a claim by targeting the person delivering it, rather than evaluating the quality of the reasons or evidence supporting or opposing it. This tactic is often used to create prejudice against an opponent and divert attention away from the substance of the argument. Ad hominem attacks are particularly common in political debates, where discussions frequently devolve into personal insults and character attacks instead of a reasoned analysis of the actual issues at hand.

A common misconception is that any criticism of a person's character automatically constitutes an ad hominem fallacy. However, this is not accurate. The fallacy specifically occurs when an attack on character is used to undermine an opponent's argument, rather than addressing the content of their reasoning or evidence.

For example, someone might say, "You shouldn't trust John's opinion on climate change because he failed science in high school." In this case, John's academic history has no bearing on whether his argument about climate change is valid or well-supported. The criticism is irrelevant to the claim being discussed.

However, there are situations where character-based criticism might be relevant and not fallacious. If the argument specifically concerns a person's integrity, behaviour, or credibility, such as in a fraud investigation or an election campaign, then discussing a person's character may legitimately contribute to the argument.

The core issue with ad hominem attacks is that the character of the person presenting an argument does not determine whether the argument itself is valid or true. For instance, it does not matter whether it was Hitler or Stalin who said that 2 + 2 = 4, the truth of the mathematical statement remains unchanged by their character or moral failings.

In short, valid arguments must stand on their own merits, regardless of who presents them. Personal attacks are irrelevant to the truth or falsity of a claim.

When encountering an ad hominem fallacy, it is important to stay focused on the argument itself. Redirect the discussion back to the premises and evidence supporting the claim, and point out that attacking the person does not address the reasoning or evidence behind their argument. It is also helpful to distinguish between legitimate character-based concerns, such as when credibility is directly relevant to the topic, and irrelevant personal attacks that serve only to distract from the issue at hand.

The ad hominem fallacy undermines meaningful discussions by shifting focus away from what is being said to who is saying it. Effective critical thinking requires evaluating arguments based on their content, evidence, and logical structure, rather than being swayed by personal attacks or irrelevant details about the individual presenting the claim. By recognising and addressing this fallacy, discussions can remain focused, productive, and rooted in reason rather than emotional or prejudicial distractions.

Figure 3.4.1 shows a few more examples of the ad hominem fallacy:

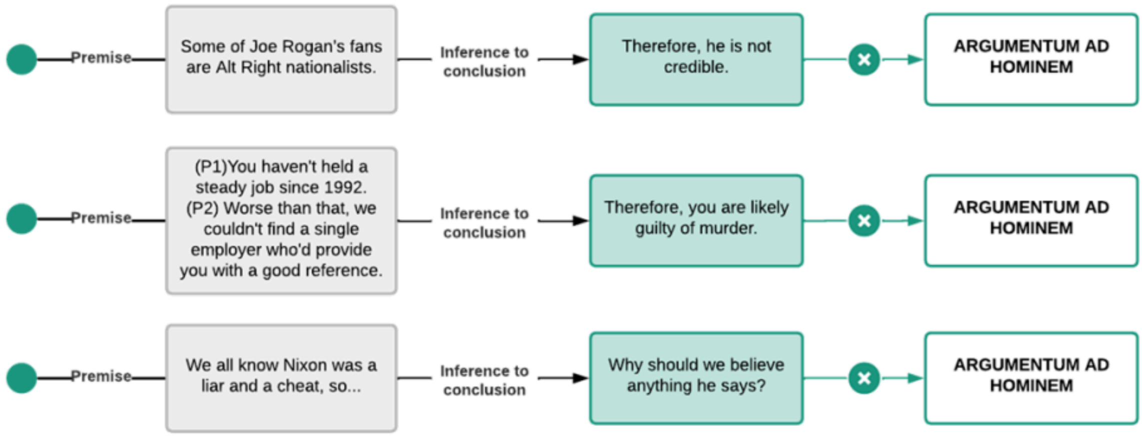


Figure 3.4.1. Additional examples of the ad hominem fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Red Herring

The red herring fallacy occurs when someone introduces an unrelated topic or irrelevant information into a discussion to distract from the original issue. The term originates from hunting, where a strongly scented fish was supposedly used to mislead hunting dogs and divert them from their intended trail. In arguments, a red herring serves a similar purpose, as it shifts attention away from the core claim or question and redirects it to something tangential or unrelated.

This fallacy is a classic example of irrelevance because it deliberately sidesteps the central argument rather than engaging with it directly. By shifting focus to a tangential topic, the person using the red herring avoids having to confront the claims, evidence, or reasoning being presented. This tactic is particularly common when someone cannot effectively counter an argument and instead changes the subject to avoid admitting weakness or conceding a point.

For example, someone might say, “We shouldn’t worry about climate change because there are so many homeless people who need our help first.” While homelessness is undeniably an important issue, it is not directly relevant to the discussion about climate change. The shift in focus serves as a distraction, preventing meaningful engagement with the original argument about climate change.

Another variation of the red herring fallacy often involves ad hominem attacks, where the focus is diverted to someone’s character or personal traits instead of their reasoning. When the accused party starts defending themselves, the original argument gets lost entirely, allowing the person who introduced the red herring to evade addressing the central claim.

The red herring fallacy is known by several other names, including “befogging the issue”, “diversion”, "ignoratio elenchi" (a Latin term meaning ignorance of refutation), "ignoring the issue", "irrelevant conclusion", and "irrelevant thesis". Despite the different labels, they all describe the same fundamental tactic: diverting attention away from the argument to avoid addressing its reasoning or evidence.

Red herrings are particularly deceptive because they disrupt the logical flow of a discussion and prevent meaningful engagement with the core argument. Instead of advancing the conversation, they rely on misdirection and distraction, leaving the original issue unresolved.

Identifying and addressing a red herring requires focus and clarity. It is essential to stay anchored to the original topic and consistently guide the discussion back to the main issue. Politely pointing out the diversion can help refocus the conversation, while asking direct, specific questions about the original claim can prevent further attempts to derail the argument.

The red herring fallacy remains a common and effective rhetorical tactic, often seen in political debates, media discussions, and heated online exchanges. Its power lies in its ability to shift attention away from challenging questions or uncomfortable evidence. Recognising when this fallacy is being used, and skilfully redirecting the conversation back to the central argument, is essential for maintaining clarity, focus, and logical consistency in any meaningful discussion.

Figure 3.4.2 shows a few more examples of the red herring fallacy:

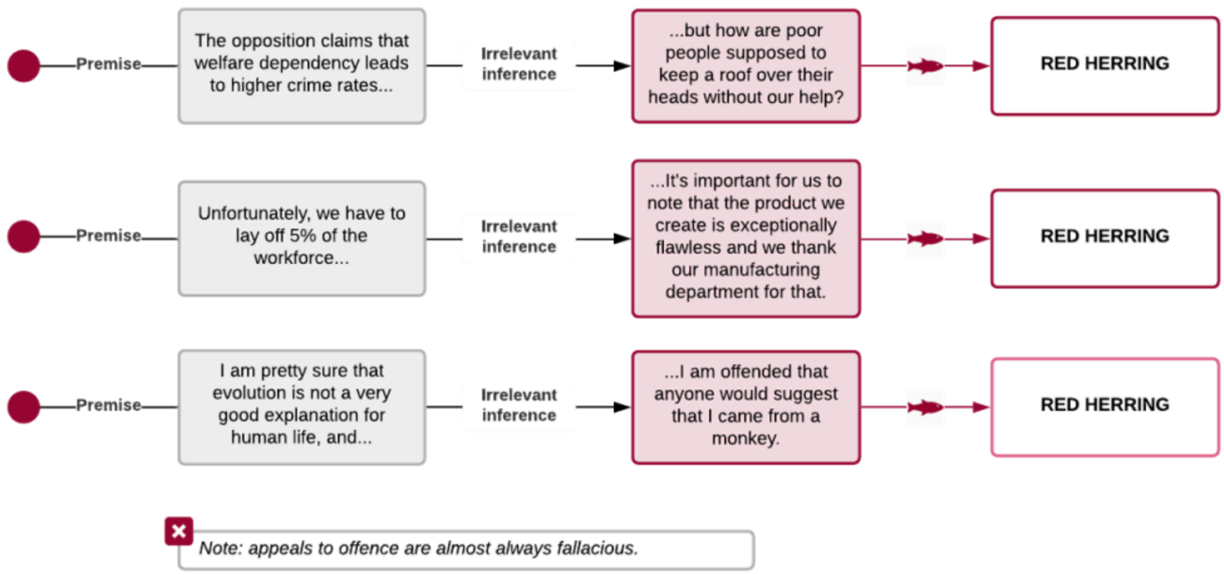


Figure 3.4.2. Additional examples of the red herring fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Tu Quoque Fallacy ("You Too")

The tu quoque fallacy, which translates from Latin as "you too", occurs when someone attempts to dismiss an argument by accusing the person making it of hypocrisy. Instead of addressing the claim or evidence, this fallacy shifts the focus onto the behaviour or character of the speaker, implying that their inconsistency invalidates their argument.

This tactic serves as both a form of ad hominem attack and a red herring, as it diverts attention away from the validity of the argument itself and redirects it toward the arguer's perceived inconsistency. The underlying assumption is that if someone does not "practice what they preach", their argument must be flawed. However, the truth or falsity of an argument is entirely independent of the person presenting it. Valid arguments rely on evidence and logical reasoning, not the personal behaviour of those who advocate for them.

For example, someone might say, "Doctor, why should I listen to your advice about quitting smoking when you're a smoker yourself?" While the doctor's smoking habit might seem like a reason to dismiss their advice, it has no bearing on the validity of the health information they are providing. The fact remains that smoking is harmful to health, regardless of whether the doctor follows their own advice.

The central issue with this fallacy is that it confuses the credibility of the speaker with the merit of their argument. While hypocrisy can undermine trust in a speaker, it does not inherently invalidate the evidence or reasoning they present. Sound arguments must always be judged on their own merits, not on whether the speaker personally adheres to the conclusions they advocate.

When encountering a tu quoque fallacy, it is essential to refocus the discussion on the argument itself. Politely point out that the truth of the claim stands independently of the speaker's actions or behaviour. Additionally, it can help to acknowledge the distraction directly, noting that accusations of hypocrisy, while perhaps worth discussing in a separate context, do not address the validity of the argument at hand. Finally, return to the core issue by evaluating the reasoning and evidence presented, rather than focusing on the behaviour or perceived inconsistency of the speaker.

The tu quoque fallacy undermines productive discussion by conflating personal behaviour with the validity of an argument. While hypocrisy can justifiably raise questions about a person's integrity or credibility, it does not automatically render their claims false. Effective reasoning requires us to separate the speaker from the argument and evaluate claims based on evidence, logic, and sound reasoning, rather than on whether the advocate perfectly embodies their own advice. Recognising and addressing this fallacy ensures that discussions remain focused, logical, and productive.

Figure 3.4.3 shows a few more examples of the tu quoque fallacy:

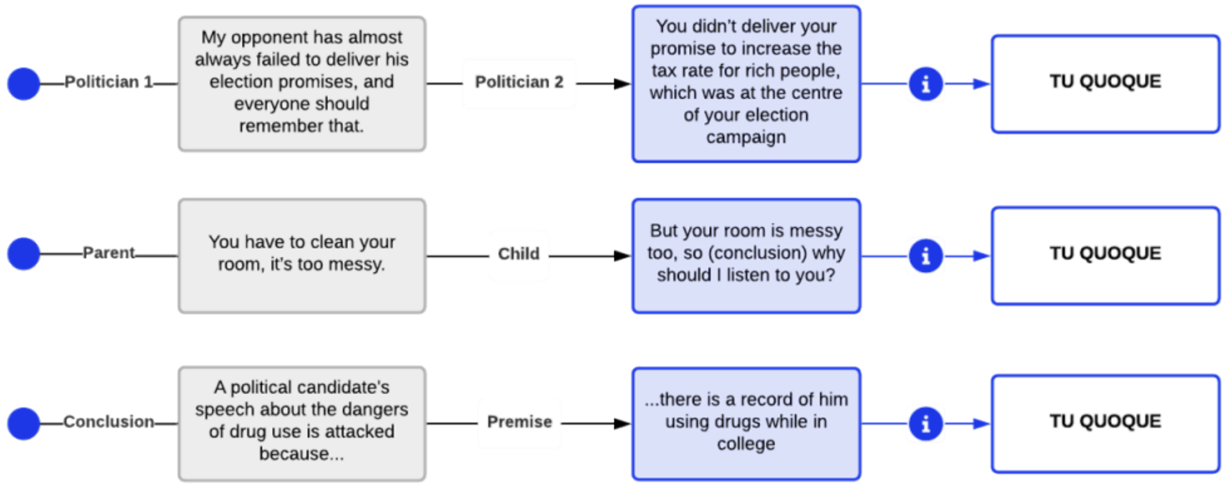


Figure 3.4.3. Additional examples of the tu quoque fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Straw Man Fallacy

The straw man fallacy occurs when someone misrepresents, oversimplifies, or exaggerates an opponent's argument to make it easier to refute. The metaphor captures this perfectly: defeating a straw man, which is a flimsy, artificially constructed version of an argument, is far easier than confronting a robust, well-supported position.

This tactic is often employed when a debater lacks sufficient reasons or evidence to directly counter their opponent's actual claims. Instead of addressing the original argument on its merits, they distort or caricature it, then proceed to dismantle this weaker, misrepresented version.

A straw man argument can take several forms. It might focus on one isolated aspect of the original claim, conveniently ignoring its broader context. Alternatively, it could exaggerate or oversimplify the opponent's position, making it appear extreme or absurd. In some cases, it involves taking statements out of context to make them sound less reasonable or easier to dismiss.

For example, someone might propose, "We should consider implementing some regulations to reduce pollution from factories." A straw man response could be, "My opponent wants to shut down all factories and destroy the economy!" In this scenario, the original argument calls for reasonable regulation, but it is deliberately distorted into an extreme, indefensible position. This exaggerated claim is far easier to argue against than the original, more nuanced suggestion.

Straw man arguments are alarmingly common in political debates, social media discussions, and contentious topics. Their emotional persuasiveness and ability to divert attention away from the core argument make them an attractive tool for those seeking to win debates rather than foster genuine understanding. Often, they provoke the opponent into defending the distorted version of their position rather than returning to their original argument.

When a straw man argument succeeds, it can mislead the audience about the opponent's actual stance, shift the focus away from valid evidence and reasoning, and make the opponent appear unreasonable or extreme.

The straw man fallacy is frequently paired with the red herring fallacy, where a deliberately distorted version of the argument serves as a distraction from the main issue. This combination allows the person using the fallacy to avoid addressing the core claims and evidence while keeping the discussion fixated on their artificial version of the opponent's argument.

Addressing a straw man fallacy begins with clarifying the original argument. It is important to clearly and directly restate the initial claim to ensure everyone understands the intended meaning. Next, it helps to point out the distortion, identifying exactly how the argument was misrepresented or exaggerated. Finally, the focus should be redirected back to the original argument and its supporting evidence, steering the discussion away from the distorted version.

The straw man fallacy undermines meaningful discussion by replacing a valid argument with a misleading caricature that is far easier to attack. Effective reasoning demands that we engage with an opponent's actual claims and evidence, not with distorted representations of them. Recognising and addressing this fallacy is essential for ensuring that discussions remain honest, focused, and productive, ultimately fostering clearer communication and better understanding.

Figure 3.4.4 shows a few more examples of the straw man fallacy:

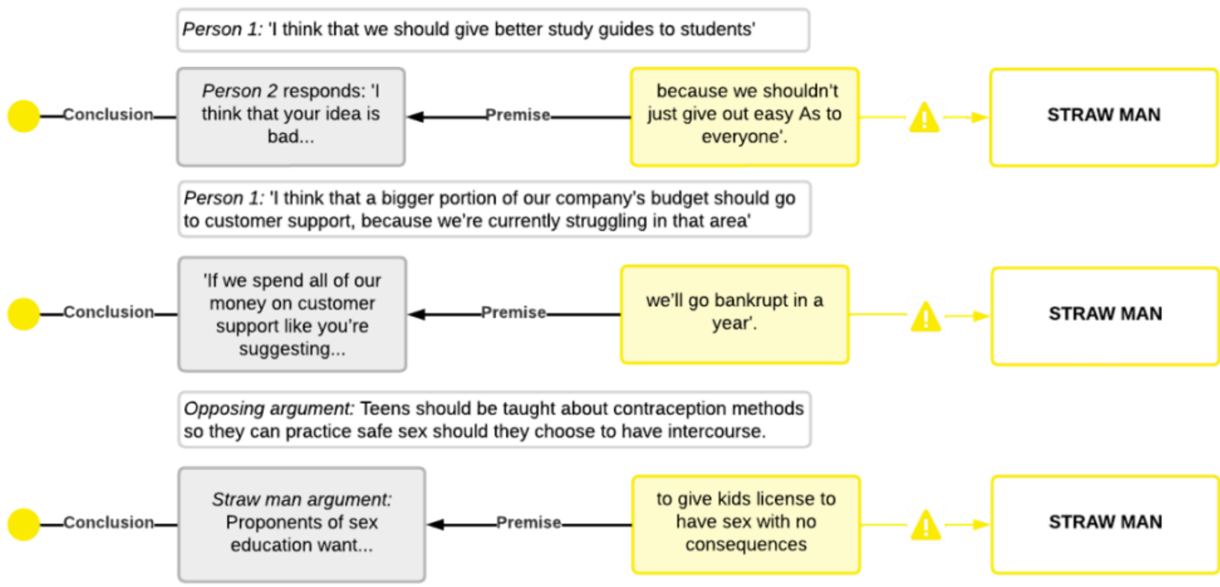


Figure 3.4.4. Additional examples of the straw man fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Appeals to Emotion

The appeal to emotion is a broad category of fallacies of irrelevance where arguments attempt to persuade through emotional manipulation rather than relying on reasoning and evidence. While emotions play an important role in human decision-making, they should never serve as a substitute for logical reasoning when evaluating the truth or validity of a claim.

These fallacies exploit our natural emotional responses, such as fear, pity, joy, or pride, to make arguments more persuasive. Emotional appeals are often highly effective rhetorical tools because humans are naturally responsive to emotional triggers. However, it is important to recognise that emotional reactions have no direct connection to whether a claim is true or false.

When someone uses emotion as a substitute for evidence, they bypass critical thinking by leveraging emotional influence. For example, an appeal to pity might sound like this: "You must pass me in this course because I've had such a hard year." Similarly, an appeal to fear could take the form of, "If we don't implement this policy, society will fall apart!" Meanwhile, an appeal to joy might claim, "Think of how happy you'll be if you buy this product!" In each case, the emotional reaction is designed to override critical analysis, making the audience more likely to accept the claim without properly evaluating its logical merits.

However, not all emotional appeals are inherently fallacious. There are situations where emotions are a valid component of reasoning, particularly in moral arguments or calls to action. For instance, in moral reasoning, emotions like compassion or outrage can provide relevant context for understanding the moral weight of an issue, such as arguments concerning human rights or environmental justice. Similarly, in motivational arguments, emotions play a crucial role in inspiring action. An appeal to urgency or responsibility, for example, might be entirely appropriate when encouraging people to donate to disaster relief efforts.

Even in these cases, however, emotional appeals must complement reasoning and evidence, not replace them. Emotion can enhance persuasion, but it must never be used to suppress facts, as ignoring evidence simply because it feels uncomfortable is intellectually dishonest. Likewise, emotional discomfort does not invalidate a logically sound argument, nor can strong emotions distort objective reality.

When encountering emotional appeals, it is essential to critically examine the argument. The first step is to identify whether the argument relies on emotional triggers instead of presenting evidence. If it does, it is important to ask for clear facts or reasoning to support the claim being made. Throughout the discussion, it is equally important to remain grounded in reasoning, acknowledging the emotional component of the argument but focusing on whether it logically holds up under scrutiny.

Emotional appeals are undeniably powerful tools for persuasion, but they become fallacious when they replace reasoning and evidence instead of working alongside them. While emotions can be relevant in moral reasoning or motivational contexts, they must always remain anchored in facts and logical analysis. Effective critical thinking requires us to acknowledge emotions without allowing them to override objective reasoning, ensuring that our conclusions are based on clear evidence and rational judgement rather than fleeting emotional responses.

Figure 3.4.5 shows a few more examples of the appeals to emotion fallacy:

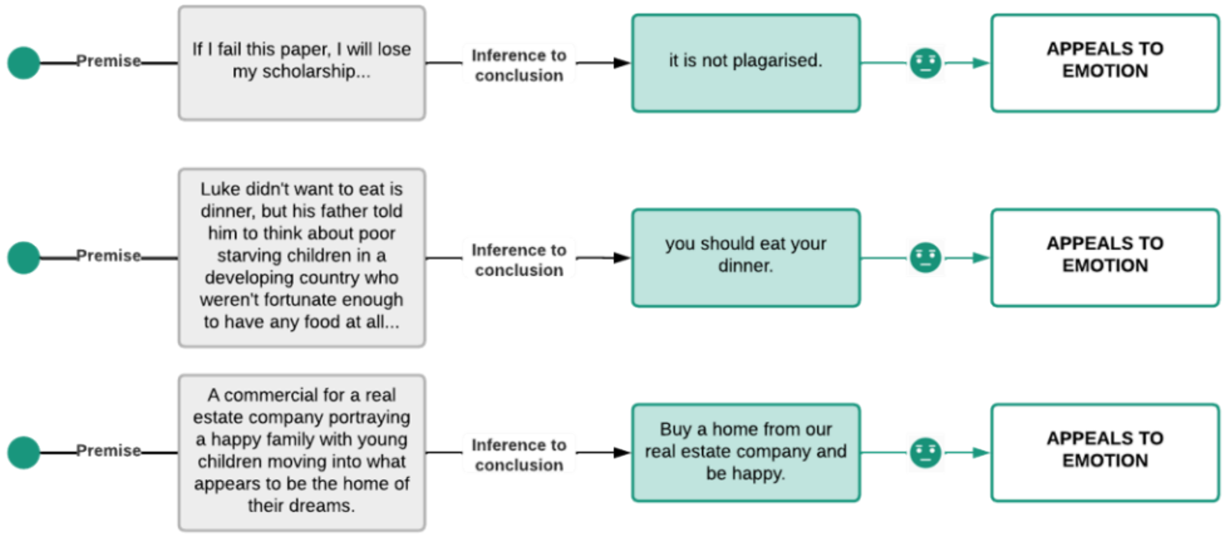


Figure 3.4.5. Additional examples of the appeals to emotion fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

---

Chapter Attribution

Content adapted, with editorial changes, from:

Reasoning critically in *Mastering thinking: Reasoning, psychology, and scientific methods* (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

3.5. FALLACIES OF AMBIGUITY

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Fallacies of ambiguity occur when unclear, vague, or misleading language is used in an argument, either intentionally or unintentionally. Unlike fallacies of insufficiency, where reasoning or evidence is incomplete, or fallacies of irrelevance, where unrelated premises are introduced, fallacies of ambiguity exploit confusion over the meaning of words or phrases to create the illusion of a valid argument.

At the heart of these fallacies lies the misuse or shifting of meanings, often without clarification. Words and phrases frequently have multiple meanings, and ambiguity arises when an argument subtly shifts between these meanings without making the transition explicit. This linguistic sleight of hand creates a false impression of coherence or support for a conclusion, even though the reasoning itself remains fundamentally flawed.

Equivocation

The equivocation fallacy occurs when a key term or phrase is used in different senses within the same argument, leading to confusion or misleading reasoning. In everyday language, it is common for words to have multiple meanings, and this is usually not an issue. However, in arguments, clarity and consistency are essential. The responsibility falls on the person presenting the argument to ensure that key terms are used consistently throughout and always refer to the same concept or idea each time they appear.

This fallacy typically occurs when a word with multiple meanings shifts subtly during the course of an argument. Sometimes, this shift is obvious and easy to spot, but in many cases, it can be surprisingly subtle and require careful attention to detail to identify.

For example, consider the following argument: "Only man is rational. No woman is a man. Therefore, no woman is rational." In this case, the word "man" is first used to mean "human beings in general", and then it shifts to mean "male humans". The conclusion relies entirely on this shift in meaning to create a faulty argument.

While examples like this one are blatant, many instances of equivocation are much more difficult to detect because the shift in meaning can be nuanced or context-dependent.

Equivocation is sometimes paired with the shifting goalposts fallacy, where the criteria for evidence or reasoning are subtly adjusted during a discussion. In such cases, someone may change the definition or expectation tied to a key term, making their argument harder to challenge.

For instance, someone might say: "Science can't explain love." When presented with studies on the biology of love, they might shift the definition of "love" from something biological and chemical to something spiritual or metaphysical. This shift in definition undermines the discussion because it redefines the original premise to avoid being addressed directly.

To address equivocation, it is essential to clarify key terms by asking for clear and consistent definitions of important words or phrases. Pay close attention to where a term seems to shift meaning midway through the argument and ensure that key terms remain consistent from premise to conclusion.

The equivocation fallacy relies on the ambiguity of language to mislead or confuse. While some cases are easy to spot, others demand careful analysis and attention to detail. Ensuring that key terms are clearly defined and consistently applied throughout an argument is vital for maintaining logical clarity and validity. By identifying and addressing subtle shifts in meaning, we can prevent this fallacy from undermining rational discourse.

Figure 3.5.1 shows a few more examples of the equivocation fallacy:

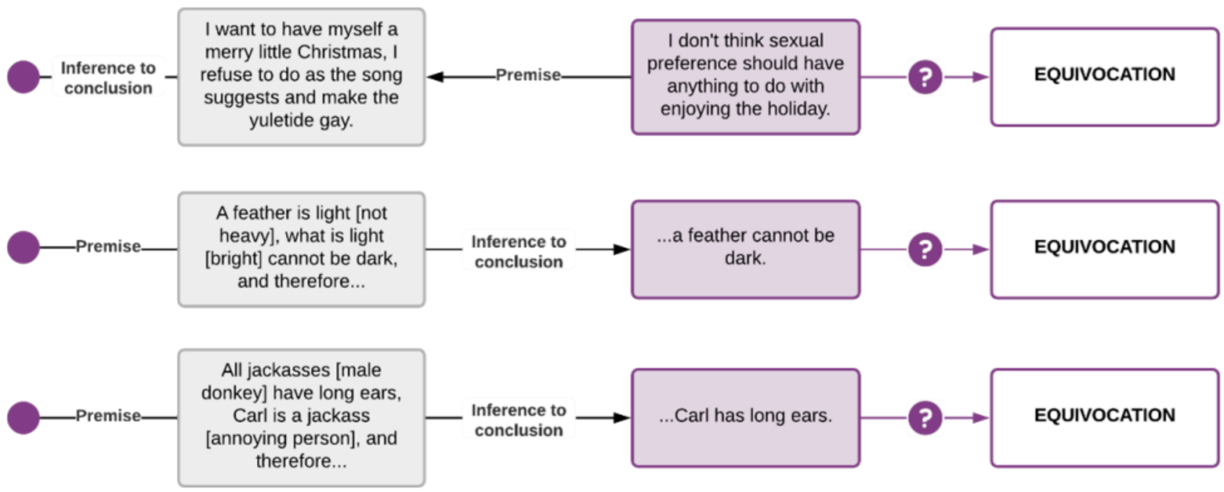


Figure 3.5.1. Additional examples of the equivocation fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Amphiboly

The amphiboly fallacy occurs when the structure or grammar of a sentence creates ambiguity, leading to multiple interpretations that can mislead or distort reasoning. Unlike equivocation, which involves shifting the meaning of a single word, amphiboly arises from poorly constructed or ambiguous phrasing in an entire sentence or phrase. This ambiguity often allows an argument to be interpreted in more than one way, sometimes leading to faulty conclusions.

In many cases, amphiboly happens unintentionally due to awkward sentence structure or misplaced modifiers, but it can also be deliberately crafted to obscure meaning and make a weak argument seem stronger. The responsibility lies with the person presenting the argument to ensure that their phrasing is clear, precise, and unambiguous.

For example, consider the sentence: "I shot an elephant in my pyjamas." The structure of this sentence leaves room for two interpretations: either the speaker was wearing pyjamas when they shot the elephant, or the elephant was somehow wearing the pyjamas. The ambiguity arises from the placement of the phrase "in my pyjamas", making it unclear which subject the phrase modifies.

In arguments, amphiboly can be more subtle but equally misleading. Take this example: "The professor said on Monday he would give a lecture on ethics." Here, it is unclear whether the professor made the statement on Monday or if the lecture will take place on Monday. This ambiguity creates room for misunderstanding and allows different interpretations to serve as a convenient escape for someone unwilling to clarify their reasoning.

Amphiboly can also intersect with other fallacies, such as red herrings or shifting goalposts, where the ambiguity is used strategically to divert attention or shift the focus of an argument. By leaving a sentence or phrase open to multiple meanings, the person using amphiboly can evade accountability or avoid addressing the central issue directly.

To address amphiboly, it is essential to clarify ambiguous phrasing by asking the speaker to restate their argument in clearer terms. Pay close attention to grammatical structure and context, and identify any points where a sentence could plausibly have more than one interpretation. In cases where the meaning remains unclear, insist on a precise explanation to ensure the argument can be properly evaluated.

The amphiboly fallacy highlights how poor grammar or sentence construction can distort reasoning and obscure the clarity of an argument. While some examples are obvious and humorous, others are far more subtle and demand careful analysis. Ensuring that sentences are well-structured and unambiguous is critical for maintaining logical consistency and clarity in arguments. By identifying and addressing amphiboly, we can prevent ambiguous phrasing from misleading discussions or undermining valid reasoning.

Figure 3.5.2 shows a few more examples of the amphiboly fallacy:

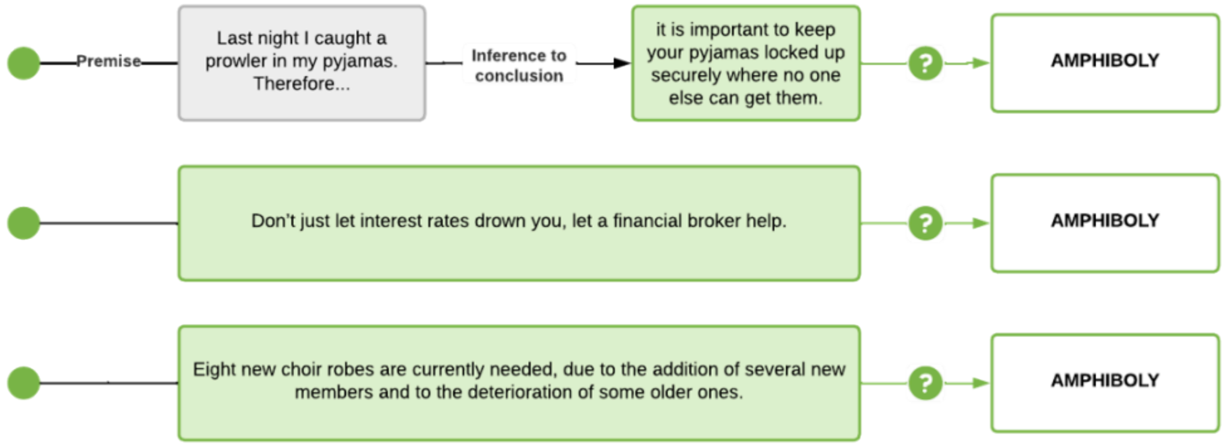


Figure 3.5.2. Additional examples of the amphiboly fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Fallacy of Composition

The fallacy of composition occurs when someone assumes that what is true of individual parts must also be true of the whole they make up. In other words, if certain components of a system or group have specific qualities, it is assumed that the entire system or group will automatically share those same qualities.

At first glance, this may seem like an obvious or trivial mistake. After all, it is common sense that a whole is not always simply the sum of its parts. However, this fallacy is far more common than it appears and often goes unnoticed in everyday conversations, marketing campaigns, and even policy discussions. Its persuasive power comes from the instinctive assumption that scaling up from individual characteristics to collective outcomes is inherently logical.

This reasoning becomes problematic because properties that exist at the level of individual components do not always transfer cleanly to the collective whole. Complex systems often exhibit emergent properties, where the group or system behaves in ways that are not directly predictable from the traits of its individual parts.

For example, someone might say, "Each player on our team is the best in their position, so our team will be the best in the league." While it might seem reasonable at first, team success depends on collaboration, strategy, and teamwork, not just individual talent. A collection of exceptional players does not automatically guarantee an exceptional team performance.

Another example is the claim, "Every brick in this building is lightweight, so the entire building must be lightweight." While each brick may indeed be light, the combined weight of thousands of bricks creates an extremely heavy structure. In both cases, the error arises from failing to recognise how properties at the individual level interact or change when scaled up to the collective level.

This fallacy becomes especially problematic when applied to social systems, economic policies, or scientific reasoning. For example, an economic policy that benefits individuals might unintentionally harm the larger economy when scaled up, due to systemic effects that don't exist at the individual level. Similarly, assuming that a successful business team can be replicated identically across different contexts ignores the unique group dynamics and environmental factors that contribute to collective success.

The fallacy of composition also shares similarities with other reasoning errors. It contrasts with the fallacy of division, where someone assumes that what is true of the whole must also be true of its parts. Additionally, it overlaps with hasty generalisation, where conclusions about an entire population are drawn from observations of a small, unrepresentative sample.

This reasoning error is sometimes referred to by other names, including the Exception Fallacy and Faulty Induction. Regardless of the terminology, the core issue remains the same: the assumption that individual traits will scale up predictably to a collective level often overlooks key systemic dynamics.

To avoid falling into this trap, it is essential to carefully analyse the relationship between the parts and the whole. Ask whether the quality in question logically scales up when applied collectively. Consider whether emergent properties, which are traits or behaviours that arise only at the collective level, might alter the expected outcome. Finally, remain sceptical of blanket assumptions that project individual characteristics onto entire systems or groups.

The fallacy of composition serves as a reminder that reasoning from parts to wholes must be approached with careful analysis and a recognition of complexity. While there are instances where such reasoning holds true, it cannot be assumed as a universal rule. Developing an awareness of this fallacy helps us avoid oversimplified conclusions and ensures more precise, logical reasoning when evaluating collective claims.

Figure 3.5.3 shows a few more examples of the composition fallacy:

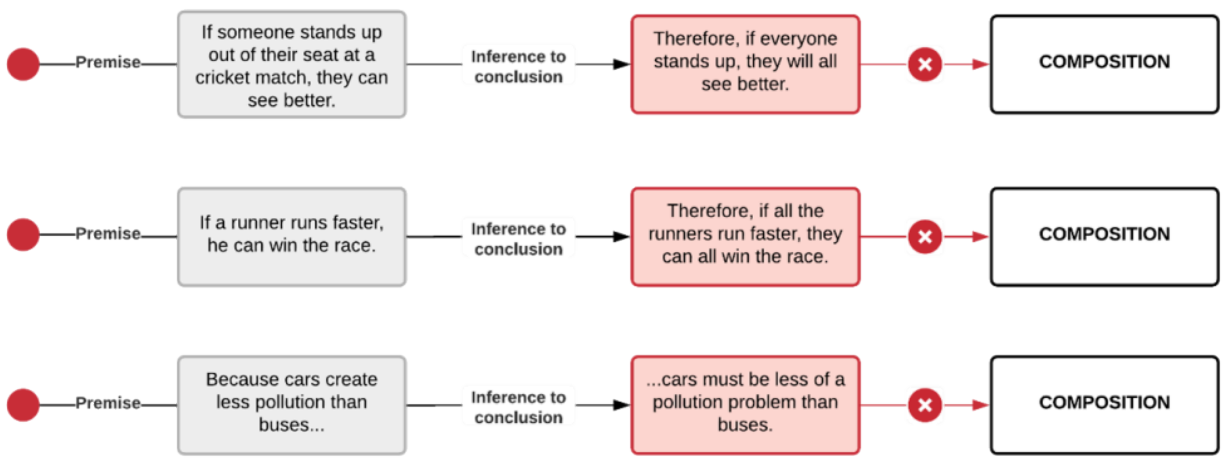


Figure 3.5.3. Additional examples of the composition fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Fallacy of Division

The fallacy of division occurs when someone assumes that what is true of a whole must also be true of its individual parts. In other words, if a group, system, or collective possesses a certain characteristic, it does not automatically follow that each member or component of that group shares the same characteristic.

This reasoning error is essentially the reverse of the fallacy of composition, where properties of individual components are incorrectly projected onto the whole. In both cases, the mistake stems from failing to recognise the distinction between collective and individual properties. While characteristics can sometimes scale up from individuals to the group, or down from the group to individuals, they often do not, especially in complex systems where emergent properties or interactions between components play a crucial role.

For example, someone might claim, "The university is prestigious, so every professor working there must also be prestigious." While the university as a whole may have an outstanding reputation, not every individual professor automatically shares that prestige. Institutional reputation depends on a collective contribution from faculty, administration, resources, and historical achievements, not solely on individual members.

Another example is the statement, "The basketball team is unbeatable this season, so every player on the team must also be unbeatable." While the team might indeed be performing exceptionally well, its success likely comes from collaboration, strategy, and group dynamics, rather than every player being exceptionally skilled on their own.

The fallacy arises because group-level characteristics do not always scale down to individual members. Outcomes or traits observed at the collective level often emerge from systemic interactions, cooperation, or structural factors, which are elements that do not necessarily translate to isolated parts.

This reasoning error becomes especially harmful when it fuels stereotypes or unjust assumptions about individuals based on group-level data. For example, one might say, "Statistically, members of Group X have lower average educational outcomes; therefore, this specific individual from Group X must also be poorly educated." This assumption is both unfair and logically flawed, as group averages cannot reliably predict individual characteristics. Data about collectives should always be interpreted with caution, especially when applied to specific cases.

The fallacy of division shares similarities with other logical missteps. It contrasts with the fallacy of composition, where individual traits are incorrectly projected onto a collective group. Additionally, it bears some resemblance to hasty generalisation, where broad conclusions are drawn from a small or unrepresentative sample. However, while hasty generalisation typically scales upward from a limited observation, the fallacy of division scales downward from a group-level observation.

This reasoning error is sometimes referred to by other names, such as False Division or Faulty Deduction. While the term "deduction" in this context can be somewhat confusing, the core issue remains the same:

assuming that group-level properties apply uniformly to individual members without sufficient justification.

To avoid committing the fallacy of division, it is essential to carefully examine the relationship between the whole and its parts. Ask whether the property in question logically transfers from the collective to the individual. Be cautious of overgeneralisations based on averages or group characteristics and always consider the context, including whether the observed property depends on systemic factors or collective dynamics rather than individual traits.

The fallacy of division serves as a reminder that group-level truths cannot always be applied to individual members without careful analysis. While some properties may indeed scale down, many do not. Recognising this fallacy helps us avoid stereotyping, challenge faulty assumptions, and analyse arguments with greater precision and fairness.

Figure 3.5.4 shows a few more examples of the division fallacy:

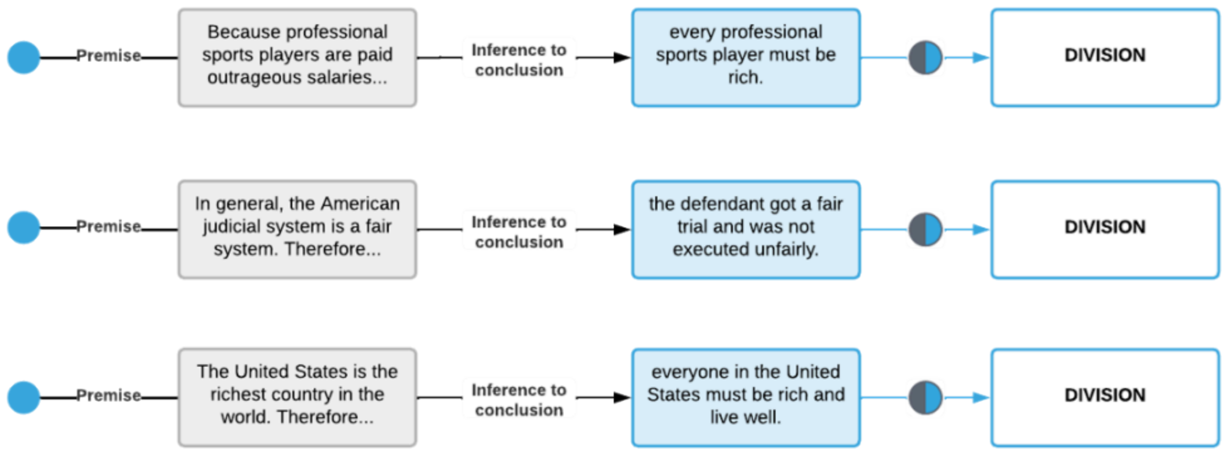


Figure 3.5.4. Additional examples of the division fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Moving the Goalposts

The "moving the goalposts" fallacy occurs when someone changes the criteria for success or acceptance in an argument after those criteria have already been met. This tactic ensures that no matter how valid or well-supported an opponent's evidence or reasoning may be, it can never fully satisfy the evolving standards.

The metaphor originates from sports, where physically moving the goalposts would make scoring a goal impossible. In debates or discussions, the effect is the same: the standards for evidence or reasoning are repeatedly adjusted, often in increasingly unreasonable ways, to ensure that the original claim remains perpetually unproven.

This fallacy often appears when someone is unwilling to concede defeat, even when their opponent has provided clear and compelling evidence. For example, a sceptic might demand, “Show me evidence of evolution happening today.” When presented with valid examples, they might respond with, “That’s not enough; show me an example of entirely new genetic information arising by random processes.” In this case, the criteria for acceptable evidence keep shifting, making it impossible to satisfy the demands. The tactic guarantees that the person moving the goalposts can always claim the evidence is insufficient, regardless of how well it meets the original request.

The issue with moving the goalposts is that it creates unfair standards that prevent meaningful resolution. First, it establishes an intellectual dishonesty, as the person employing the tactic demonstrates a lack of genuine openness to evidence or reasoning. Second, it leads to endless demands where the discussion becomes circular, with no clear way to reach a conclusion.

This fallacy thrives in ambiguous discussions, especially when terms like “proof”, “evidence”, or “compelling reason” are left undefined. Without clarity on what constitutes acceptable evidence, one party can continuously shift expectations, making productive dialogue impossible.

To avoid falling into this trap, it is important to define clear standards at the outset of a discussion. Both parties should agree on what counts as acceptable evidence or reasoning before proceeding. Once these standards are established, they should remain consistent throughout the discussion. If someone begins altering their requirements, it is essential to politely call out the inconsistency and redirect the conversation back to the original agreement.

For example, a clear agreement might sound like this: “If I can show you two independently verifiable examples of evolution happening today, will you accept that as evidence?” Establishing this kind of standard sets clear expectations and reduces the likelihood of goalpost-shifting later in the discussion.

The “moving the goalposts” fallacy is sometimes referred to by other names, including Raising the Bar, Shifting Sands, Gravity Game, and Argument by Demanding Impossible Perfection. Each of these terms highlights the same core issue: unfairly altering the conditions for acceptance after the discussion has already begun.

Ultimately, the “moving the goalposts” fallacy undermines productive dialogue by preventing arguments from reaching a fair and meaningful resolution. Avoiding this fallacy requires clear agreements about what constitutes valid evidence, a commitment to consistent standards, and a willingness from all parties to engage in good-faith reasoning. By maintaining these principles, discussions can remain focused, fair, and intellectually honest.

Figure 3.5.5 shows a few more examples of the moving the goalposts fallacy:

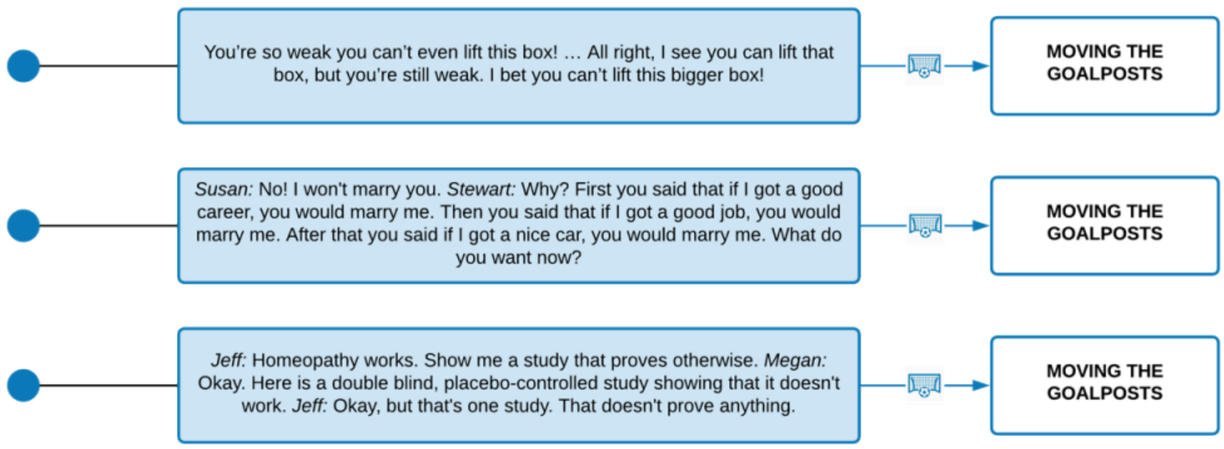


Figure 3.5.5. Additional examples of the moving the goalposts fallacy by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

---

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

3.6. COGNITIVE BIASES

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Many of the fallacies we have discussed are so persuasive because they tap into our cognitive biases, which are built-in tendencies that make us vulnerable to flawed reasoning, incomplete evidence, ambiguous language, and irrelevant distractions. These biases can significantly interfere with our ability to think critically and rationally, which is precisely the focus of this chapter. One of the primary reasons we often fail to reason effectively is our blindness to these cognitive biases and the subtle ways they shape our judgement.

Cognitive biases are systematic patterns of deviation from rational judgement, where the framing or context of information distorts how we perceive, evaluate, and decide. As we have already explored, humans rely on mental filters and shortcuts (heuristics) to process information quickly and efficiently. These mental shortcuts evolved not to guarantee accuracy or correctness, but to help us make fast and generally useful decisions in survival-oriented contexts. However, while heuristics can simplify complex decision-making, they also introduce errors and distortions, making our reasoning more prone to mistakes.

While informal fallacies and cognitive biases both lead to flawed reasoning, they differ in their origins and how they manifest. Informal fallacies are errors in the structure or content of an argument, often resulting from poor reasoning, misrepresentation, or misuse of evidence. They are typically identifiable within the framework of a specific argument. In contrast, cognitive biases are deeply ingrained psychological tendencies, which are systematic patterns of thought and judgement that operate subconsciously. Fallacies are usually found in external arguments, while biases are internal habits of thought that influence how we interpret and construct those arguments in the first place.

At this point, it might feel like this chapter is turning into something of a 'listicle', which is a term often used for articles that are structured as lists rather than fully developed discussions. That is because, just like informal fallacies, cognitive biases are commonly presented in categorised lists. A quick online search will return titles like "The Top 10 Cognitive Biases You Need to Know" or "5 Cognitive Biases That Shape Your Thinking". However, the goal here is not to overwhelm you with an exhaustive catalogue but to introduce a foundational set of key biases that will enable you to recognise and understand others more easily.

In cognitive psychology, particularly in the context of cognitive therapy, the term 'cognitive bias' takes on a more specific meaning. It refers to habitual patterns of distorted thinking, often referred to as cognitive distortions, that contribute to and exacerbate emotional distress, anxiety, and depression. If you are studying psychology, you will likely encounter these concepts in more detail. For the purposes of this chapter, however, we are using the term ‘cognitive bias’ in a broader sense, referring to general thinking habits that can lead to reasoning errors.

One of the most intriguing and ironic aspects of cognitive biases is our tendency to easily spot them in others while remaining oblivious to them in ourselves. This bias, in itself, is a cognitive bias. It is incredibly common to notice flawed reasoning, selective interpretation, or emotional decision-making in someone else while overlooking the same patterns in our own thought processes.

For this reason, it is crucial to remain open to feedback from others when they point out potential biases in your reasoning. If you are like most people, and you are, you likely have a few cognitive biases that you are unaware of. Recognising and acknowledging them is the first step toward more accurate, balanced, and reflective reasoning.

The Dunning-Kruger Effect

The Dunning-Kruger Effect describes a fascinating and counterintuitive phenomenon: the less someone knows about a topic, the more confident they are in their knowledge. You have likely encountered this effect in everyday life, even if you were not aware it had a name.

This cognitive bias was first formally described by psychologists David Dunning and Justin Kruger in their influential 1999 study titled “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments.” In essence, the Dunning-Kruger Effect highlights a troubling truth: the less knowledge or skill someone has in a given area, the less capable they are of recognising their own limitations.

At its core, the effect shows that ignorance breeds overconfidence. People with limited understanding of a subject often overestimate their expertise because they lack the self-awareness to recognise their own gaps in knowledge. On the other hand, individuals who possess genuine expertise tend to be more cautious and humble in their assessments, often underestimating their own competence because they are keenly aware of what they do not know.

Even Charles Darwin observed a version of this effect long before it was given a formal name, writing that “Ignorance more frequently begets confidence than does knowledge.”

The Dunning-Kruger Effect serves as a powerful reminder of the importance of intellectual humility and the need for self-awareness in evaluating our own abilities and knowledge. Understanding this bias can help us approach unfamiliar topics with a more balanced perspective and encourage more thoughtful self-assessment.

Confirmation Bias

Confirmation bias is one of the most widespread and influential cognitive biases. At its core, confirmation bias refers to our tendency to seek out, interpret, remember, and give more weight to information that supports our existing beliefs, while ignoring or downplaying evidence that contradicts them.

This bias stems from our deep psychological attachment to being right. Changing our minds is uncomfortable, mentally taxing, and often feels like admitting failure. As a result, we become emotionally invested in our existing beliefs and subconsciously filter the world to reinforce them.

In the Information Age, confirmation bias has become an even bigger problem. With unlimited access to information online, it is incredibly easy to find articles, studies, or opinions that align with whatever belief we already hold. A quick Google search can provide endless "evidence" to validate almost any position, no matter how flawed or incorrect. This creates an echo chamber effect, where we shield ourselves from opposing viewpoints and avoid confronting the possibility that we might be wrong.

But confirmation bias does not just affect what information we seek out, it also influences how we process and remember evidence. When presented with two pieces of evidence, one that supports our belief and one that challenges it, we are far more likely to scrutinise and dismiss the contradictory evidence while readily accepting the supporting evidence. Furthermore, we are more likely to remember the confirming evidence and forget or distort the contradicting information over time.

Understanding confirmation bias is essential for critical thinking because it reminds us to approach evidence and opposing viewpoints with intellectual humility and an open mind. Overcoming this bias requires a conscious effort to actively seek out disconfirming evidence, question our assumptions, and evaluate all evidence with equal scrutiny, regardless of whether it aligns with our preexisting beliefs.

Self-Serving Bias

Self-serving bias refers to our natural tendency to attribute positive outcomes to our own actions or character while blaming negative outcomes on external factors. In essence, we protect our self-esteem by taking credit for successes and shifting blame for failures onto something, or someone, outside of ourselves.

This bias serves an important psychological function: it helps us maintain a positive self-image and avoid feelings of guilt, shame, or inadequacy. When something goes well, we are quick to attribute the success to our skills, intelligence, or effort. However, when something goes wrong, we are equally quick to point to bad luck, other people's mistakes, or uncontrollable circumstances as the cause.

A classic example can be seen in the aftermath of a car accident. Both drivers involved are far more likely to blame the other party, even if they themselves contributed to the collision. Similarly, in academic or professional settings, people often credit their hard work and intelligence for their successes but blame unfair teachers, bad bosses, or difficult circumstances for their failures.

Interestingly, self-serving bias does not operate the same way in everyone. Individuals with low self-esteem or depression may actually experience a reversed self-serving bias. In these cases, people are more likely to blame themselves for negative outcomes and attribute positive outcomes to luck or external factors, further reinforcing their negative self-perception.

Understanding the self-serving bias is essential for developing self-awareness and accountability. Recognising when we are falling into this pattern can help us take responsibility for our actions, learn from failures, and grow from our experiences rather than defaulting to protective but ultimately unproductive mental habits.

The Curse of Knowledge and Hindsight Bias

Knowledge, while valuable, comes with its own set of drawbacks. One of these is the curse of knowledge, sometimes referred to as the curse of expertise. This bias occurs when someone who is well-versed in a topic fails to recognise how much less others may know about it. Experts, such as lecturers or seasoned professionals, often assume that their audience shares their foundational understanding. This assumption can lead to poor communication, misunderstandings, and unrealistic expectations about how others will interpret or act on information.

The problem arises because once we have fully integrated a piece of knowledge into our understanding of the world, it becomes intuitive and seemingly obvious. Explaining something to someone without that foundational knowledge suddenly feels far more difficult than we realise. This creates barriers to teaching, collaboration, and even predicting how others might respond in certain situations.

Closely related is hindsight bias, which deals not with knowledge but with events and outcomes. After something significant happens, it feels inevitable in retrospect, even if it was not obvious beforehand. This bias convinces us that we "knew it all along" or that the outcome was clearly predictable. However, this false sense of foresight overlooks the uncertainty and complexity present before the event occurred.

A common example of hindsight bias can be found in studying historical events. When analysing the events leading up to World War I, for instance, it is tempting to wonder how experts at the time failed to see the impending crisis. With the clarity of hindsight, every detail appears to have been a clear warning sign, even though those living through the events were navigating ambiguity and incomplete information.

Both the curse of knowledge and hindsight bias highlight how our perspective on knowledge and events changes once we have additional information. Being aware of these biases can help us communicate more effectively, remain humble about what we "knew" beforehand, and approach complex situations with a clearer understanding of uncertainty and perspective.

Optimism and Pessimism Bias

Humans have a notoriously poor grasp of probability, and optimism and pessimism biases are clear examples of this shortcoming. These biases influence how we perceive the likelihood of positive or negative outcomes, and they often lead us to make flawed judgements about future events. Our ability to accurately assess probabilities is heavily influenced by factors such as our personality, mood, and the nature of the situation we are evaluating.

Optimism bias leads us to overestimate the likelihood of positive outcomes and underestimate the risk of negative ones. For example, many university students surveyed believe they are less likely to experience negative life events, such as divorce or alcohol addiction, compared to their peers. At the same time, they tend to overestimate their chances of positive outcomes, like owning a home or living past the age of 80. These skewed perceptions are often reinforced by confirmation bias, which makes it easy for us to focus on evidence that supports our optimistic expectations while dismissing contradictory information.

Conversely, pessimism bias causes people to overestimate the likelihood of negative outcomes and underestimate positive possibilities. While optimism and pessimism biases seem like opposites, they can actually coexist within the same person, depending on the specific scenario. For example, someone might feel overly optimistic about their career prospects while simultaneously being overly pessimistic about their health outcomes.

It is also worth noting that pessimism bias can be more pronounced in individuals with mental health conditions, such as depression. These individuals may consistently interpret future events with an exaggerated sense of risk or inevitability of failure.

Ultimately, both optimism and pessimism biases reveal how subjective and unreliable our assessments of probability can be. Recognising these biases helps us approach future planning with greater realism and balance, making us more aware of our tendency to lean too far in either direction when predicting outcomes.

The Sunk Cost Fallacy

The sunk cost fallacy occurs when we continue investing time, money, or effort into something simply because we have already invested so much, even when it no longer makes sense to do so. A sunk cost refers to any expense, whether financial, emotional, or in terms of time, that has already been incurred and cannot be recovered. Rational decision-making tells us that these past costs should not influence our future choices, but our emotions and psychological biases often override this logic.

This fallacy leads us to overvalue past investments while undervaluing future or ongoing costs. Essentially, the more we have invested in a decision, the harder it becomes to walk away, even when the most logical choice would be to stop. For example, someone might overeat at a buffet because they feel they need to "get their money's worth", even if they are uncomfortably full. Similarly, a business might continue funding a failing project simply because they have already poured significant resources into it, rather than redirecting those resources to something more promising.

At its core, the sunk cost fallacy exploits our emotional attachment to past investments. We are naturally resistant to the idea of "wasting" what we have already put in, even if persisting results in further losses. The belief that we must "see it through" creates a powerful psychological pull, making it feel like abandoning the effort is a failure rather than a wise strategic choice.

Recognising the sunk cost fallacy involves a shift in perspective; learning to evaluate decisions based on their future value and potential outcomes rather than past investments. By focusing on what can still be gained (or avoided) moving forward, rather than what has already been lost, we can make more rational and effective decisions.

Negativity Bias

Humans have a natural tendency to focus more on negative experiences and emotions than on positive ones, even when their intensity is the same. This phenomenon, known as negativity bias, means that negative events have a greater psychological impact on us than positive ones of equal significance.

For example, we are far more likely to fixate on an insult or mistake than to dwell on a compliment or success. Even if the praise and criticism are equally strong, the criticism tends to linger in our minds, while the praise fades more quickly. This imbalance is not just a quirk of personality; it is deeply rooted in how our brains process emotional experiences. Negative emotions and events are registered more intensely, are more easily recalled, and tend to dominate our thought patterns.

This bias serves an evolutionary purpose. Historically, paying close attention to potential threats or negative events increased our chances of survival. However, in modern contexts, this bias can distort our perspective, making us overly focused on problems, setbacks, or critical feedback while overlooking positive experiences and successes.

It is important to note that negativity bias is distinct from pessimism bias. While negativity bias focuses on how we process past and present events, pessimism bias is about our expectations for future events. Understanding this distinction helps clarify how these biases influence our thoughts and emotions in different contexts.

Recognising negativity bias allows us to be more intentional about balancing our focus on positive experiences and not letting negative events dominate our mental space. By consciously acknowledging and celebrating positive outcomes, we can counterbalance this natural tendency and develop a more even-handed perspective on our experiences.

The Backfire Effect

We like to think of ourselves as rational beings, ready to adjust our beliefs when presented with new facts and evidence. However, reality tells a different story, one that becomes painfully clear after five minutes on social media. Instead of welcoming evidence that contradicts our views, we often respond by digging in our heels even deeper.

The backfire effect describes this counterintuitive reaction. When confronted with evidence that challenges deeply held beliefs, instead of reconsidering our stance, we often become even more committed to our original position. It is as if admitting we were wrong is so uncomfortable, so threatening to our sense of self, that we would rather reject reason and evidence entirely.

Rather than softening our stance, new information can feel like an attack on our identity, triggering a defensive response where we double down on our original belief, as though preparing for a long and stubborn stand-off. This effect is especially common with beliefs tied to our identity, values, or worldview.

The irony of the backfire effect is that it often strengthens the very beliefs it seeks to challenge, making productive conversations around controversial topics incredibly difficult. It is a psychological bunker mentality: when challenged, we fortify our mental defences rather than opening the gates for reflection and growth.

And like many cognitive biases, while it is easy to spot this behaviour in others, it is notoriously difficult to notice in ourselves. Recognising the backfire effect in our own thinking requires humility, self-awareness, and a willingness to sit with the discomfort of being wrong, which is a task far easier said than done. Understanding this bias can help us approach difficult conversations with more patience, empathy, and a focus on collaboration rather than confrontation.

The Fundamental Attribution Error

The fundamental attribution error highlights a common imbalance in how we explain behaviour, both our own and others'. When judging our own actions, we tend to blame external circumstances, while when judging others, we often attribute their behaviour to their character or personality.

For example, if we accidentally cut someone off in traffic, we might excuse ourselves by saying, "I didn't see them" or "I was in a hurry." In other words, we justify our actions by pointing to the situation we were in. However, if someone else cuts us off, we are far less generous in our interpretation. Instead of considering that they might be rushing to an emergency or simply did not see us, we are more likely to think, "What a careless jerk!"

This bias occurs because it is easier to observe others' behaviour than the situational factors influencing them. When it comes to our own actions, we have a deeper understanding of our intentions, pressures, and constraints. But with others, we only see the outcome of their actions, not the invisible situational context behind them.

This tendency to favour dispositional (personality-based) explanations over situational ones is also known as the correspondence bias or attribution effect.

Recognising this bias is crucial because it affects our judgement, empathy, and relationships. By reminding ourselves that everyone operates within their own set of circumstances, just as we do, we can approach others' behaviour with more understanding and less immediate judgement.

In-Group Bias

Humans have a natural tendency to categorise the world into social groups, which is how we make sense of complex social dynamics. However, this inclination often leads to in-group bias, also known as in-group favouritism. This bias refers to our tendency to favour people we perceive as part of our own group, viewing them and the group as a whole, more positively than those outside it.

At first glance, this might not seem like a bias. After all, if we did not view our group positively, we would probably just join a different one. Whether it is supporting a sports team, identifying with a cultural group, or aligning with a political party, our group affiliations are deeply tied to our sense of identity and self-esteem.

What is particularly striking about in-group bias is how easily it can emerge. Studies have shown that people start displaying favouritism even when they are randomly assigned to completely meaningless groups, with no shared history or meaningful connection. In these experiments, participants often favour their new group members in everything from distributing rewards to forming opinions, even though the group itself was created arbitrarily and holds no real significance.

This bias is not inherently malicious; it is a byproduct of our natural desire to belong and feel valued. Our group memberships help shape our self-identity and reinforce our sense of self-worth. However, unchecked in-group bias can lead to unfair treatment of others, reinforce stereotypes, and create unnecessary divisions between groups.

Being aware of in-group bias allows us to reflect on our assumptions and judgements, encouraging us to evaluate others based on their individual merits rather than group affiliations. In doing so, we can move towards more balanced and fair interactions across different social groups.

The Forer Effect (also known as The Barnum Effect)

The Forer Effect, sometimes called the Barnum Effect, refers to our tendency to believe vague and general personality descriptions are uniquely tailored to us. This bias explains why people often find horoscopes, personality tests, or fortune-telling surprisingly accurate, even when the descriptions are so broad they could apply to almost anyone.

For example, a horoscope might describe someone as "strong-willed yet sensitive, often drawn to creative pursuits and deeply caring for loved ones." While these traits could easily apply to a wide range of people, individuals often interpret them as being specifically about themselves, especially if they align with their self-image.

This effect works because people tend to focus on details that feel personally meaningful while overlooking the generic nature of the description. Essentially, we are drawn to self-relevance, even in statements designed to be universally relatable.

A playful way to test the Forer Effect is to pretend you are a different zodiac sign when talking to someone who strongly believes in astrology. You will likely find they quickly identify traits from that sign in your behaviour or personality. And when you eventually reveal your true sign, they might simply brush it off with a comment like, "Oh, well, that sign is known for being deceptive!"

This interaction also highlights how confirmation bias and the backfire effect can reinforce these beliefs. Once someone feels their personality aligns with a description, they become resistant to evidence suggesting otherwise.

The Forer Effect reminds us how easily we can fall into the trap of seeing patterns and personal relevance in vague statements. Recognising this bias can help us approach generalised claims, whether in horoscopes, personality quizzes, or marketing materials, with a more critical and discerning mindset.

Chapter Attribution

3.7. SUMMARY

By Marc Chao

Summary

Understanding logical, physical, and technological possibilities provides a robust framework for evaluating what is achievable within various contexts. Logical possibility ensures coherence by adhering to unchanging laws of logic, such as the law of non-contradiction, forming the foundation for sound reasoning. Physical possibility is governed by natural laws, like Newton's laws of motion and thermodynamics, which distinguish feasible scenarios from those defying empirical reality. Technological possibility, dynamic and ever evolving, reflects humanity's ability to innovate within logical and physical constraints, exemplified by advancements such as space exploration and computing. Together, these categories offer a comprehensive lens for analysing problems, testing arguments, and fostering responsible innovation rooted in reality.

Fallacies of insufficiency arise when arguments fail to provide adequate evidence or reasoning to justify their conclusions, leaving them reliant on weak or unjustified assumptions. Examples include hasty generalisations, which draw broad conclusions from limited or biased samples, and post hoc fallacies, which mistakenly attribute causation based on sequential events without proof. Slippery slope arguments predict exaggerated outcomes without substantiating each step, while faulty analogies rely on superficial or irrelevant similarities. Appeals to unqualified authority and ignorance substitute credible evidence with misplaced reliance on authority or the absence of disproof. Circular reasoning assumes the conclusion within its premises, and false dichotomies oversimplify complex issues into extreme, binary choices. Identifying these fallacies enhances critical thinking by encouraging rigorous evaluation of evidence and ensuring arguments remain logical, well-supported, and free from unwarranted assumptions.

Fallacies of irrelevance distract from the central claim by introducing unrelated premises rather than addressing the argument directly. These include the ad hominem fallacy, where personal attacks discredit an argument, and red herrings, which divert attention to irrelevant topics. The tu quoque fallacy deflects criticism by highlighting perceived hypocrisy, while the straw man

fallacy misrepresents an opponent's position to make it easier to refute. Emotional appeals, another common type, rely on manipulation rather than evidence to persuade, though emotions can play a role in moral or motivational reasoning when used alongside logic. Addressing these fallacies helps maintain focus on the argument's core, ensuring discussions remain logical, relevant, and productive.

Fallacies of ambiguity exploit unclear or misleading language, creating the illusion of valid reasoning by introducing confusion over meanings or phrasing. Equivocation occurs when a word shifts meaning within an argument, while amphiboly arises from ambiguous grammar or structure that leads to multiple interpretations. The fallacy of composition assumes that what is true of individual parts applies to the whole, and the fallacy of division projects group-level traits onto individuals. Moving the goalposts shifts evidence standards, rendering arguments perpetually unprovable. These fallacies undermine logical reasoning by distorting clarity or consistency, emphasising the importance of precise language, fair standards, and careful analysis in maintaining valid arguments. Addressing these errors fosters clearer, more effective discussions.

Cognitive biases are ingrained tendencies that distort judgement and reasoning, leaving us vulnerable to flawed arguments, misinterpretations, and irrational decisions. These biases operate subconsciously, shaping how we process information, evaluate evidence, and form beliefs. Examples include the Dunning-Kruger Effect, where ignorance fosters overconfidence, and confirmation bias, which leads us to favour information that aligns with existing beliefs while disregarding contradictions. The sunk cost fallacy traps us into continuing commitments based on past investments, and the fundamental attribution error attributes others' actions to character rather than context. Emotional and social biases, such as negativity bias, optimism bias, and in-group bias, further skew perceptions. The Forer Effect demonstrates how we find personal relevance in vague statements, while the backfire effect reveals resistance to changing beliefs when challenged. Recognising these biases enhances critical thinking by fostering self-awareness, intellectual humility, and balanced decision-making.

CHAPTER 4: PERCEIVING AND BELIEVING

How do we make sense of the world? Why do two people witnessing the same event often recall it so differently? And how do our deeply held beliefs shape the way we perceive reality? In this chapter, we delve into the intricate relationship between perception and belief systems, exploring how sensory information is filtered, interpreted, and often reshaped by our prior knowledge, experiences, and biases.

Perception forms the foundation of how we navigate and understand our environment. It involves not only the raw data gathered through our senses but also the complex processes our brain uses to organise and interpret this information. However, perception is far from a flawless or static mechanism. It evolves as we interact with new stimuli, influenced by factors such as attention, culture, and personal beliefs. This chapter begins by examining the dynamic nature of perception, including the roles of bottom-up (data-driven) and top-down (concept-driven) processes, and the fascinating interplay between the two.

From there, we explore how perception is not always aligned with objective reality. Through phenomena like optical illusions, inattentional blindness, and the cocktail party effect, we uncover how our sensory systems can deceive us. These examples serve as a reminder that what we perceive is not the world as it is, but the world as our brain constructs it.

The chapter also introduces belief systems, which are the interconnected webs of ideas that shape how we interpret our experiences and navigate the world. Our beliefs are both influenced by and influence our perceptions, creating a dynamic feedback loop. Through illustrative experiments and theoretical insights, we examine how beliefs act as filters for sensory input and how they can, in turn, be reinforced by what we perceive.

Central to this exploration is the recognition of the biases and heuristics that underpin human cognition. These mental shortcuts simplify the overwhelming complexity of the world but often come at the cost of accuracy, leading to misinterpretations and errors. By understanding how biases like confirmation bias and heuristics like chunking operate, we gain tools to critically evaluate our thought processes and improve decision-making.

Finally, this chapter highlights the importance of adopting critical thinking attitudes, such as modesty, openness, and intellectual courage, toward our perceptions and beliefs. Recognising the limitations of our sensory systems and the fallibility of our mental models allows us to question assumptions, embrace uncertainty, and remain flexible in the face of new evidence.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand the concept of perception:** Define perception as the process of interpreting sensory information and explain how it influences our understanding of the world.
- **Differentiate between bottom-up and top-down processing:** Identify the characteristics of bottom-up (data-driven) and top-down (concept-driven) processing and understand how they interact to shape perception.
- **Explain the role of attention in perception:** Describe how selective and divided attention influence what we perceive and provide examples such as the cocktail party effect and inattentional blindness.
- **Explore the influence of biases and heuristics on perception:** Recognise how cognitive shortcuts like heuristics and biases, including confirmation bias, shape our interpretation of sensory data and decision-making.
- **Analyse the impact of priming on perception and behaviour:** Explain the concept of priming, its unconscious influence on perception and behaviour, and provide examples of its effects in social and cognitive contexts.
- **Recognise the interaction between beliefs and sensations:** Describe how beliefs act as filters for sensory input and create a feedback loop that shapes both perception and belief systems.
- **Evaluate the role of cultural and personal factors in perception:** Explain how cultural backgrounds, personal experiences, and implicit biases influence perception and highlight their implications for empathy and understanding.
- **Develop critical thinking approaches to beliefs:** Cultivate attitudes such as modesty, openness, and intellectual courage toward beliefs, enabling a more objective and adaptive approach to understanding and interpreting the world.

4.1. PERCEPTION

By Judith Rafferty, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Have you ever listened to two people recall an incident and found their stories so different that you wondered if they were even talking about the same event? This striking divergence often stems from differences in sensory perception and the many factors that shape and influence how people perceive events. Sensory perception provides a useful starting point for understanding these differences.

In this section, we use the term perception to refer to the experiences that result from the stimulation of our senses and the process of making meaning from those experiences. Perception is not static. Instead, it evolves and changes based on the information we receive. While perception may often seem automatic, it is actually a complex and dynamic process that supports our actions. Imagine how you would perceive the world without your senses, such as sight, touch, smell, taste, or hearing. It would be impossible to navigate or understand your surroundings.

To introduce the concept of perception, watch the video by CrashCourse: Perceiving is Believing: Crash Course Psychology #7 up to 3:51.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=1225#oembed-1

In the video, Hank Green describes perception as "the top-down way our brains organise and interpret information and put it into context" (we will explore the term "top-down" in more detail shortly). Every brain is unique, shaped by a combination of genetic makeup and environmental factors during development. This means that how we organise, interpret, and contextualise information is heavily influenced by our nature, experiences, emotions, social environments, and cultures. Acknowledging our own "lens" and biases, meaning how we perceive and assign meaning to incoming information, is essential for developing strong critical thinking skills and appreciating the diverse perspectives people bring to discussions and problem-solving situations.

It is also important to recognise that our perceptions are not always accurate and can be easily fooled. For example, watch this optical illusion video by eChalk, which demonstrates how our visual perception can deceive us [2:05].

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=1225#oembed-2

As Hank Green states in Perceiving is Believing: "Sometimes, what you see is not actually what you get." Similarly, Feldman Barrett (2017) explains in her book How Emotions Are Made that "you see what your brain believes" (p. 78). Feldman Barrett refers to this phenomenon as "affective realism", which she explores in greater detail. Understanding that our beliefs influence our perceptions can also help us see how these beliefs shape our emotions.

Bottom-Up and Top-Down Processing

Bottom-up processing is often described as "data-driven" because it begins with our sensory receptors. These receptors gather sensory information from the environment and send signals to the brain, which then processes the data to construct a perception. This type of processing relies solely on the sensory input itself.

In contrast, top-down processing occurs when we interpret sensory information based on our prior experiences, knowledge, and expectations. This is often referred to as concept-driven or schema-driven processing because it uses pre-existing mental frameworks to make sense of what we perceive.

Take a look at the following image in Figure 4.1.1. What do you see? Spend a few moments trying to make sense of the black blobs in the picture.

Figure 4.1.1. Illusion from The Intelligent Eye by Gregory (1970).

If you have never seen this image before, you will likely continue seeing random black blobs, no matter how long you look. This demonstrates how bottom-up processing, such as sensory stimulation alone, can be insufficient to create an accurate perception. However, once you know what the picture depicts (a Dalmatian sniffing the ground in front of a tree), your perception changes dramatically. As Feldman Barrett (2017) explains, "once you have been cured of your experimental blindness" (p. 26), your brain groups certain blobs as part of the Dalmatian and others as shadows in the background.

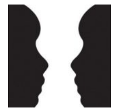
This shift occurs because neurons in your visual cortex adjust their firing, creating connections and outlines that are not physically present in the image. Essentially, your brain constructs the Dalmatian based on your new understanding of the image, a classic example of top-down processing. From this point onward, you will likely recognise the Dalmatian in the picture every time you see it, thanks to your prior knowledge.

The Gestalt Approach

The Gestalt approach provides another perspective on how we understand and interpret perception, offering principles that are highly relevant to critical thinking. Table 4.1.1 below lists the key principles of the Gestalt approach.

One particularly interesting principle is "closure", which refers to our tendency to fill in missing or incomplete information to create a cohesive image. This principle extends beyond visual perception and influences how we process and interpret information in general. In critical thinking, this tendency can lead individuals to fill in gaps in knowledge or reasoning with assumptions that align with their pre-existing beliefs or perspectives.

Table 4.1.1. Summary of Gestalt principles of form perception

Principle	Description	Example	Image
Figure-ground relationship	We structure input so that we always see a figure (image) against a ground (background).	At right, you may see a vase or you may see two faces, but in either case, you will organise the image as a figure against a ground.	
Similarity	Stimuli that are similar to each other tend to be grouped together.	You are more likely to see three similar columns among the *XYX* characters at right than you are to see four rows.	
Proximity	We tend to group nearby figures together.	Do you see four or eight images at right? Principles of proximity suggest that you might see only four.	
Continuity	We tend to perceive stimuli in smooth, continuous ways rather than in more discontinuous ways.	At right, most people see a line of dots that moves from the lower left to the upper right, rather than a line that moves from the left and then suddenly turns down. The principle of continuity leads us to see most lines as following the smoothest possible path.	
Closure	We tend to fill in gaps in an incomplete image to create a complete, whole object.	Closure leads us to see a single spherical object at right rather than a set of unrelated cones.	

Gestalt Principles and Perceptual Hypotheses

Gestalt theorists explain that pattern perception, which is our ability to differentiate figures and shapes, follows specific principles like closure, proximity, and similarity. These principles guide how we organise sensory information into meaningful wholes. However, our perception does not always match reality.

Gestalt theorists argue that perception is guided by perceptual hypotheses, where our brains make educated guesses when interpreting sensory information. These hypotheses are shaped by our personalities, experiences, and expectations and help us generate a perceptual set (a mental framework for interpreting sensory input). For example, research has shown that verbal priming can bias how people interpret ambiguous figures, demonstrating how our expectations influence what we perceive (Goolkasian & Woodbury, 2010, as cited in Stevens & Stamp).

The Depths of Perception: Bias, Prejudice, and Cultural Factors

Perception is a complex process, influenced by sensations as well as personal experiences, biases, prejudices, and cultural backgrounds. These factors can lead to significant differences in perception between individuals. Research has revealed that implicit biases, such as racial prejudice, can significantly impact perception.

For example:

- Studies have shown that non-Black participants are quicker to identify objects as weapons, and more likely to misidentify non-weapons as weapons, when those objects are paired with images of Black individuals (Payne, 2001; Payne, Shimizu, & Jacoby, 2005, as cited in Stevens & Stamp, 2020).
- Similarly, in video game studies, White participants made decisions to shoot armed targets more quickly when the targets were Black, compared to non-Black targets (Correll, Park, Judd, & Wittenbrink, 2002; Correll, Urland, & Ito, 2006, as cited in Stevens & Stamp, 2020).

This research has troubling implications, particularly when considering the many high-profile cases in recent decades involving young people of colour being killed by individuals who believed, often mistakenly, that these unarmed individuals posed a threat or were armed. Understanding how biases and cultural factors shape perception is essential for addressing these issues and fostering critical thinking.

Naïve Realism

Our perceptual systems function much like an augmented reality system, honed by millions of years of evolution. Long before modern technology created games like Pokémon GO, our brains were augmenting reality to help us interpret and navigate the world. While this system is a remarkable evolutionary achievement, it can lead to errors when our perceptions do not align with what is actually "out there".

One clear and entertaining way to observe this is through perceptual illusions. For example, in the centre of Figure 4.1.2, you might see a downward-pointing white triangle. However, this triangle does not exist, it is your brain's top-down processing creating an illusion. Similarly, in Figure 4.1.3, you may see slanted horizontal lines, but there are none; the lines are perfectly straight and parallel. Once again, your perceptual system has augmented reality, creating an experience that does not match objective reality.

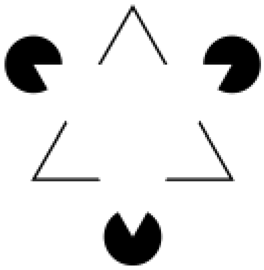


Figure 4.1.2. "The Kanizsa triangle optical illusion" by Fibonacci is used under a CC BY-SA 3.0 licence

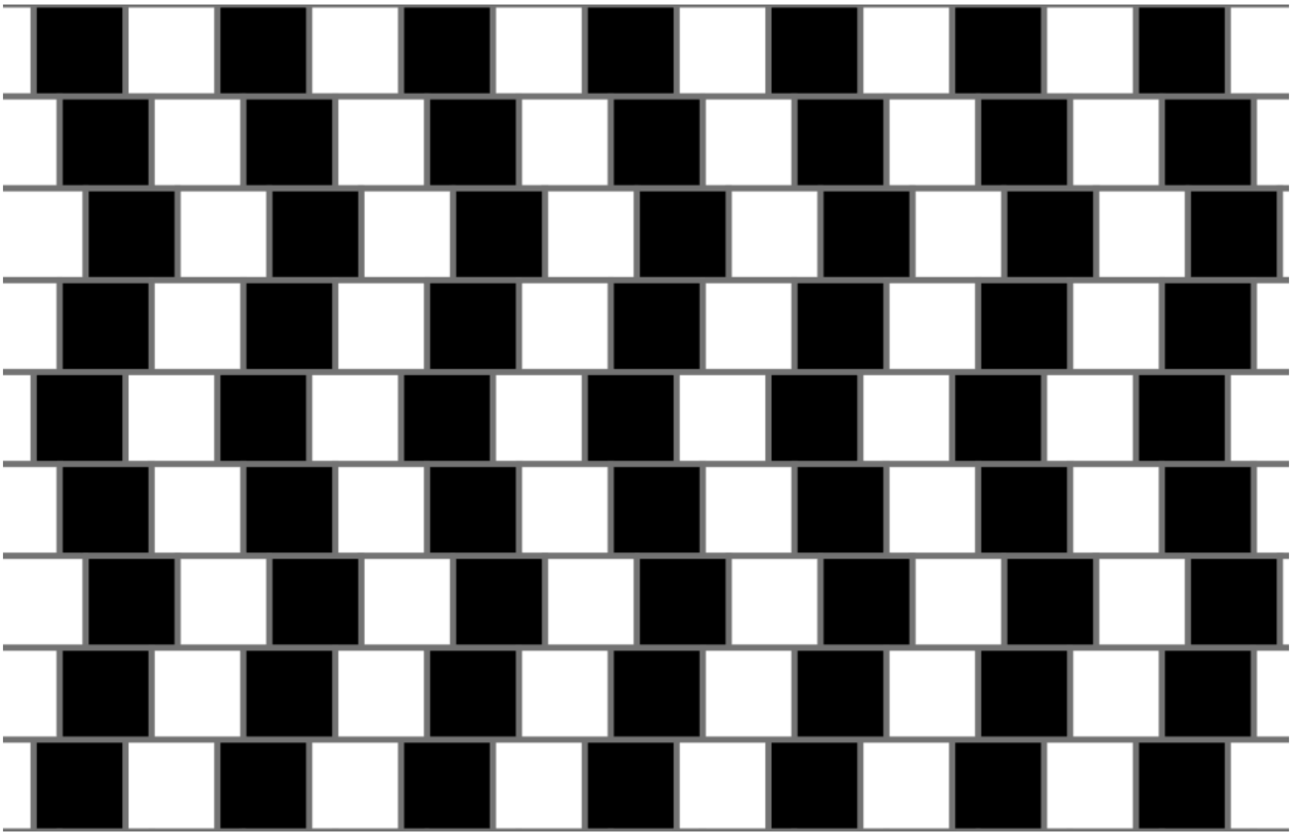


Figure 4.1.3. "The Café Wall optical illusion" by Fibonacci is used under a CC BY-SA 3.0 licence

Without a healthy dose of science, psychology, and philosophy, many people have a commonsense view of perception that is overly trusting. We tend to place too much faith in our senses, despite their numerous flaws. This perspective is sometimes compared to that of a trusting lover who repeatedly falls for the same lies, only to be betrayed again. In philosophy and psychology, a version of this overly trusting perspective is known as naïve realism.

What is Naïve Realism?

Naïve realism is the belief that everyday objects, like desks, trees, or rain, exist as they appear to us (the "realism" part) and that our senses give us direct and accurate access to these objects (the "naïve" part). According to this view, as we move through the world and interact with objects, they make contact with our sensory systems (like our eye retinas and eardrums), and we perceive them exactly as they are.

This commonsense perspective is practical for navigating daily life. After all, it would be exhausting and unproductive to constantly doubt what your senses show you. However, naïve realism has its limitations. Our sensory systems are far from infallible and leave us vulnerable to misperceptions and errors.

References

Feldman Barrett, L. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

Stevens, L. & Stamp, J. (2020). *Introduction to psychology & neuroscience* (2nd ed.). Dalhousie University. https://caul-cbua.pressbooks.pub/intropsychneuro/

Chapter Attribution

Content adapted, with editorial changes, from:

Neuroscience, psychology and conflict management (2024) by Judith Rafferty, James Cook University, is used under a CC BY-NC licence.

4.2. SENSES AND SIGNALS

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

---

The term "signal" often carries the implication of "information". For example, when you pick up a radio or cellular signal, you are receiving meaningful data. In discussions about information, "signal" is frequently contrasted with "noise", which refers to irrelevant or non-informative signals. Essentially, anything that conveys information we are interested in can be considered a signal. Conversely, randomness does not convey meaningful information, making it inherently noisy. This is because random processes, such as a coin toss, cannot be predicted or correlated with anything else, and therefore carry no meaning or useful information.

To illustrate this, imagine stargazing in search of a particular constellation. In this scenario, the signal is the specific light emitted by the stars you are trying to observe, while the noise is the light coming from other, irrelevant sources. Whether something is considered signal or noise depends on our goals and what we are trying to detect.

Our senses and brains are remarkably adept at detecting, filtering, and processing signals amidst noise. Much of this work occurs without conscious effort, allowing us to focus on what is most relevant. However, this process is not infallible. It is influenced by various biases, which can shape how we perceive and interpret signals.

The Sensory Apparatus: How We Detect Signals

An overview of how our body detects and processes signals provides a helpful foundation for understanding sensation and perception. Our sensory system includes a network of receptors designed to detect various signals. Some of these receptors are external, located on the surface of our bodies, such as our eyes, ears, nose, and mouth, which function similarly to the peripherals of a computer, like a mouse or keyboard. Others are internal, like pain receptors, which can be compared to a computer's internal heat sensors. External receptors detect forms of incoming energy (light, sound, smell, taste), converting them into electrical signals that are transmitted to the brain. At this stage, the process of "sensation" occurs, though no conscious "perception" or experience has yet taken place.

Raw sensations, on their own, are often not very useful until the brain processes them. However, there are exceptions, such as the rapid "fight or flight" response triggered by the detection of threats. This automatic system bypasses conscious perception, producing immediate physical reactions to prepare the body for action. For instance, sensations can initiate bodily responses to danger, like increased heart rate or adrenaline release, even before we consciously feel fear. This recognition comes only after noticing our physical reaction.

While this explanation of "bottom-up" sensation and perception is widely accepted, it is somewhat oversimplified. It assumes that sensory signals are limited, clear-cut, and processed passively by the brain. In reality, neither assumption holds true. Our environment is saturated with countless signals, along with a significant amount of irrelevant noise. As the 19th-century psychologist William James famously observed, the world of a newborn is a "blooming, buzzing confusion". This description applies to the flood of signals overwhelming our senses at any stage of life.

To make sense of this chaos, cognitive processes are necessary to guide the interpretation of sensory input in a "top-down" manner. For example, while a newborn's cognitive faculties gradually develop to impose order on sensory input, an adult deprived of these top-down influences would experience the same overwhelming confusion. Thus, perception is not a passive reception of sensory data but an active process shaped by both incoming sensations and the brain's interpretive mechanisms.

Our Senses in Bottom-Up and Top-Down Processing

Cognitive psychology and perception theories distinguish between two key processes: top-down and bottom-up. Top-down processing originates from the brain and mind, actively shaping what we perceive based on prior knowledge, expectations, and context. In contrast, bottom-up processing builds perceptions from raw sensory input, starting with the information gathered by our senses. Both processes work together as we interpret and understand the world around us.

Perceptual illusions occur because they exploit the dominance of top-down processing over bottom-up sensations, leading to perceptions that do not correspond to reality. For instance, in Figure 4.2.1 (known as 'Ebbinghaus' or 'Titchener Circles'), top-down influences relying on context make it challenging to recognise that the two orange circles are actually the same size.

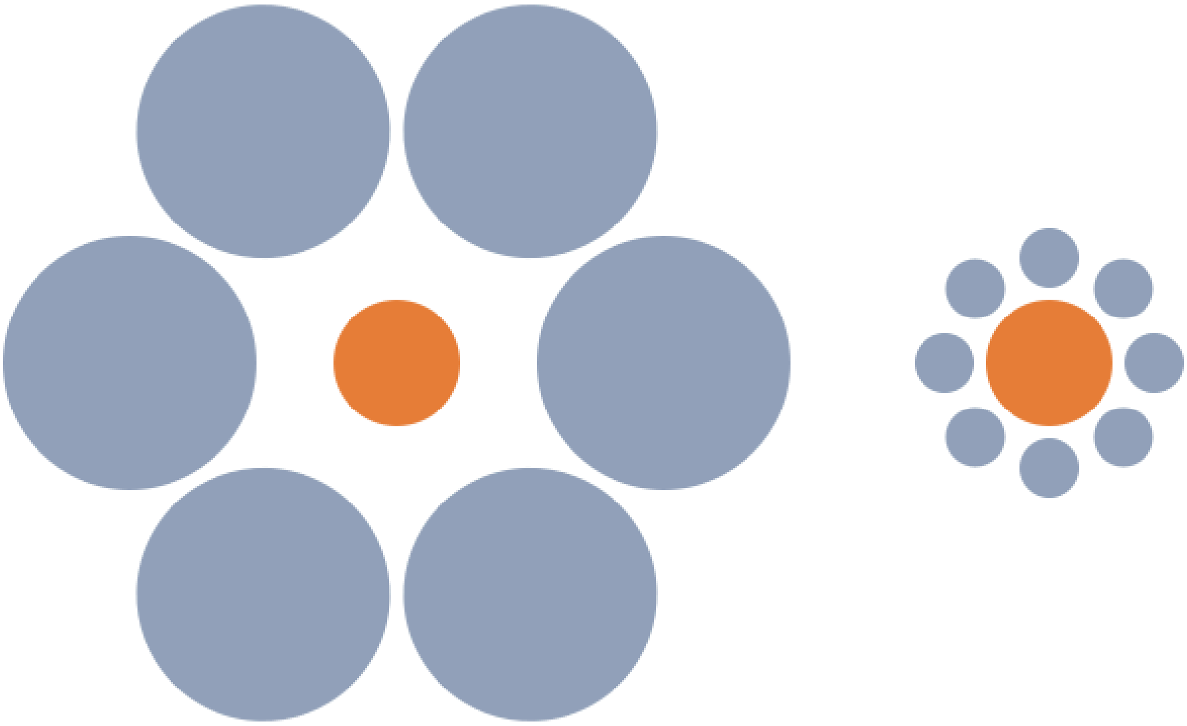


Figure 4.2.1. "Ebbinghaus' Titchener Circles" by Phrood~commonswiki is in the Public Domain, CC0

The distinction between top-down and bottom-up processes is a valuable framework for understanding how we perceive and interpret the world. Both processes are essential, as our perceptions result from a combination of bottom-up sensory input and top-down cognitive influences (refer to Figure 4.2.2).

The first role of our sensory perception system is to direct and discriminate incoming information. The influence of top-down processes on sensation should not be seen as a drawback; it is, in fact, essential. Without it, meaningful perception would be almost impossible. This aligns with William James's idea of the "blooming buzzing confusion", which describes an unfiltered sensory world. For instance, the vast number of photons striking the retina at any given moment would be overwhelming if our brains were not actively focusing on specific signals while filtering out others.

One example of the necessity of top-down processing is the cocktail party effect, which enables us to focus on a single voice in a noisy environment filled with competing sounds. The fact that we can filter out countless voices to concentrate on one person speaking is a remarkable feat of perception.

In this light, perception is best understood as a creative act. Our brain and mind do not passively receive raw sensory input; instead, they actively filter and transform it into meaningful perceptions. These perceptions may or may not accurately reflect reality. Importantly, the goal of our sensory systems is not to produce completely accurate representations of the world. Rather, they have evolved to create perceptions that enhance our survival. For the brain, survival value is a higher priority than accuracy, and this shapes the way we perceive and interpret the world around us.

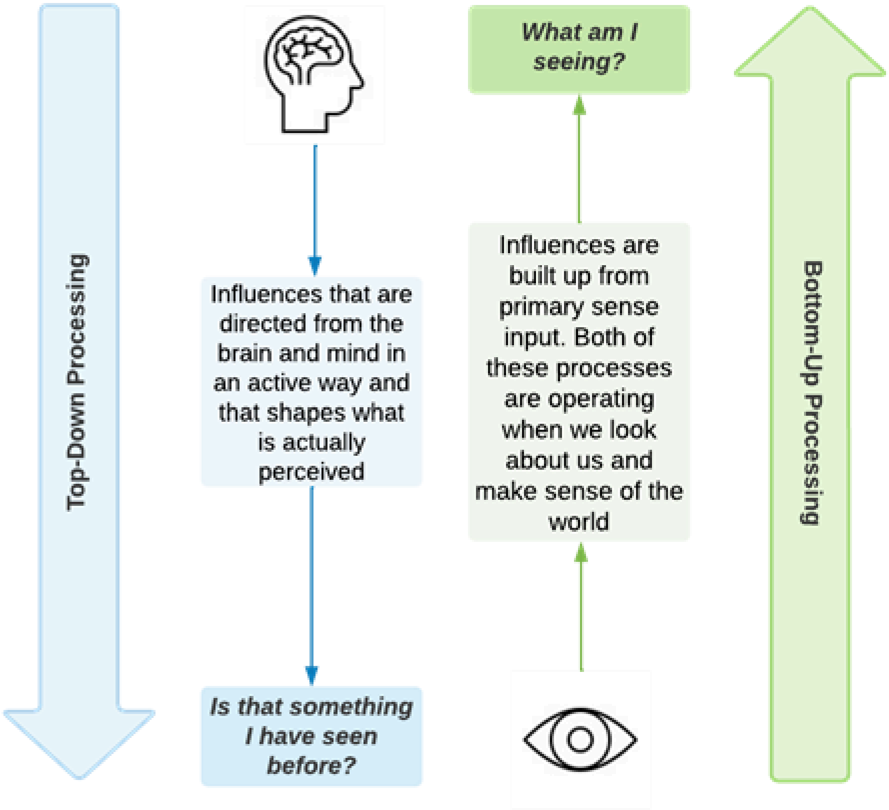


Figure 4.2.2. "Top-down and bottom-up processing" by Michael Ireland is used under a CC BY-SA licence

An example of how top-down processing can go wrong occurs in certain mental illnesses. Many mental health conditions affect how we perceive the world around us. This is especially noticeable in psychotic disorders, where people experience delusions (false beliefs) and hallucinations (auditory or visual experiences that are not real). However, this phenomenon is also present in other mental health conditions, though less obvious. For example, individuals with anxiety disorders tend to have an increased sensitivity to and focus on signals they perceive as potentially threatening (Kindt & Van Den Hout, 2001). Additionally, these individuals often struggle to distinguish between neutral and genuinely threatening stimuli (Laufer et al., 2016). In these cases, the brain's top-down processing actively shapes and alters primary sensory input, influencing how sensations are perceived.

These top-down influences are not limited to mental illness. Many other factors also affect how we perceive and interpret the world. One important group of influences aims to improve the speed and efficiency of our cognitive processing. These influences are called heuristics, which are mental shortcuts that help us quickly process information and make decisions.

Heuristics

Heuristics are strategies that serve as shortcuts for processing information and making decisions. They help simplify complex tasks by acting as "rules of thumb", enabling us to quickly perceive and interpret information while reducing the mental effort or resources required. However, heuristics come with a significant drawback: they are prone to errors. These shortcuts often aim for "good enough" results rather than perfectly accurate or optimal outcomes. As a result, while heuristics save time and mental energy, they can also lead to biases and mistakes.

One example of a heuristic is the fight-or-flight response, which enhances survival by triggering a rapid physiological reaction to perceived threats. This response can protect us in dangerous situations, but it can also lead to false alarms, such as reacting to a coiled hose as if it were a snake. Another example of heuristics in action occurs when we instinctively pay closer attention to our valuables around someone who fits a stereotypical profile, such as a person with dreadlocks, shabby clothes, or face tattoos. This snap judgement is based on a heuristic that may associate outward appearance with criminality, even though such features are unrelated to a person's behaviour. Reflecting more critically might lead us to realise the flaw in this automatic assumption.

For further reading and examples of heuristics, the [Verywell Mind website] offers additional insights and resources.

If you are unsure about how much you rely on heuristics in your daily thinking, consider the following problems. Most people answer these incorrectly due to heuristic-based thinking. While the questions themselves are not particularly difficult, your reliance on heuristics can make it easy to overlook the correct answers. Once you see the solutions, you might find yourself thinking, "How did I miss that?"

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=35#h5p-11

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://jcu.pressbooks.pub/critical-thinking-psychology/?p=35#h5p-12

Here is a final example: a logic puzzle created by Peter Cathcart Wason in 1966. In this puzzle, you are presented with four cards, each with a number on one side and a colour on the other.

The rules are as follows:

- The numbers can be either odd or even.
- The colours can be either blue or green.
- The claim to test is: “If a card shows an even number on one face, then its opposite face is blue.”

Your task is to determine which two cards need to be flipped over to test this claim properly.

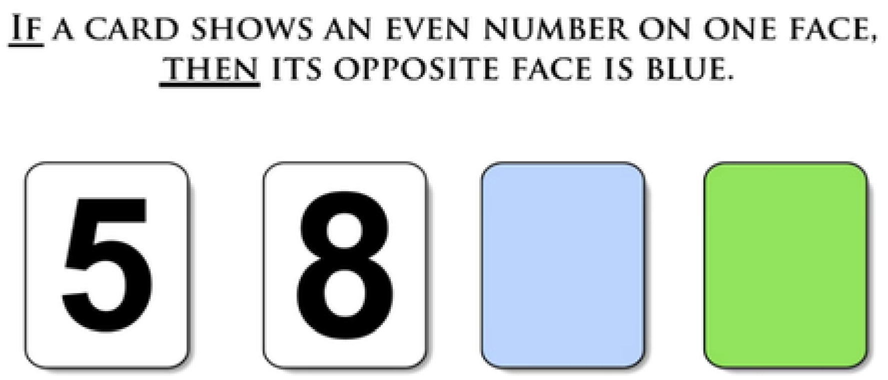


Figure 4.2.3. “Card task” Adapted by Michael Ireland (from Life of Riley) is used under a CC BY-SA licence

A helpful tip is that if you got it wrong, you were likely focusing on confirming the statement rather than attempting to falsify it.

We would not want to completely eliminate heuristics from our lives, as they are essential for managing the overwhelming amount of cognitive processing and decision-making we face each day. However, to use heuristics effectively, we must understand their influence on our thinking. Heuristics can be thought of as mental shortcuts that “bias” us toward interpreting and understanding things in specific ways. While the term bias often has a negative connotation, in this context, it simply refers to a tendency to process information in a certain way, which can be either positive or negative.

Biases have evolved because they provide significant benefits. For example, overreacting to a coiled hose (mistaking it for a snake) has clear survival value. However, some biases can be harmful. These are the ones that distort our thinking, skew our judgement, and lead us away from the truth. When people refer to “bias”, they are usually talking about this negative type.

A common example of a heuristic is a stereotype, which allows us to make quick, though often biased, judgements about someone. While stereotypes may have had evolutionary advantages for survival, they often lead to unfair or inaccurate assumptions.

In critical thinking, our primary concern is with heuristics that we call cognitive biases. These are systematic patterns of biased thinking, and in many cases, the terms “heuristics” and “cognitive biases” can be used interchangeably. As we explore this further, you will see how these concepts shape our perceptions and decisions.

Signal Detection & Transformation

As mentioned earlier, our bodies have many sense organs that detect inputs both from inside and outside. These include the five basic senses most people are familiar with: sight (vision), hearing (audition), smell (olfaction), taste (gustation), and touch (somatosensation). However, we actually possess far more than just these five senses, and some scientists suggest we may have as many as 20.

Among these additional senses are ones we use daily without much thought. For instance, proprioception allows us to sense the position of our body parts without looking at them (you can test this by closing your eyes and touching your nose). Similarly, equilibrioception (sense of balance) and thermoception (sense of temperature) are essential but often taken for granted. Other senses include chronoception (sense of time passing) and kinaesthesia (sense of movement).

You might wonder: What does this have to do with critical thinking? The point is that all of these senses are imperfect tools, sensitive to specific types of input but prone to errors. Understanding the limitations and vulnerabilities of our sensory systems is key to developing the right mindset for critical thinking. This includes cultivating humility about our beliefs and maintaining vigilance to account for how these limitations influence how we form and revise our views.

The impressions our senses register provide raw data, but it is our brains that do the heavy lifting, such as filtering, interpreting, and transforming this input into our picture of the world. Importantly, science has shown that this sensory representation is neither complete nor entirely accurate. Our senses are bombarded with billions of signals, and without the brain filtering and organising this data, we would be overwhelmed and unable to function. These filters are essential for making sense of the world. Without them, we would perceive everything as chaotic and incomprehensible, much like staring at those old Magic Eye pictures (also called stereograms or autostereograms) until a clear image emerges.

For instance, Figure 4.2.4 contains a hidden image of the African continent on a globe. I placed it in the image myself using a fun website that creates stereograms. See if you can find it.

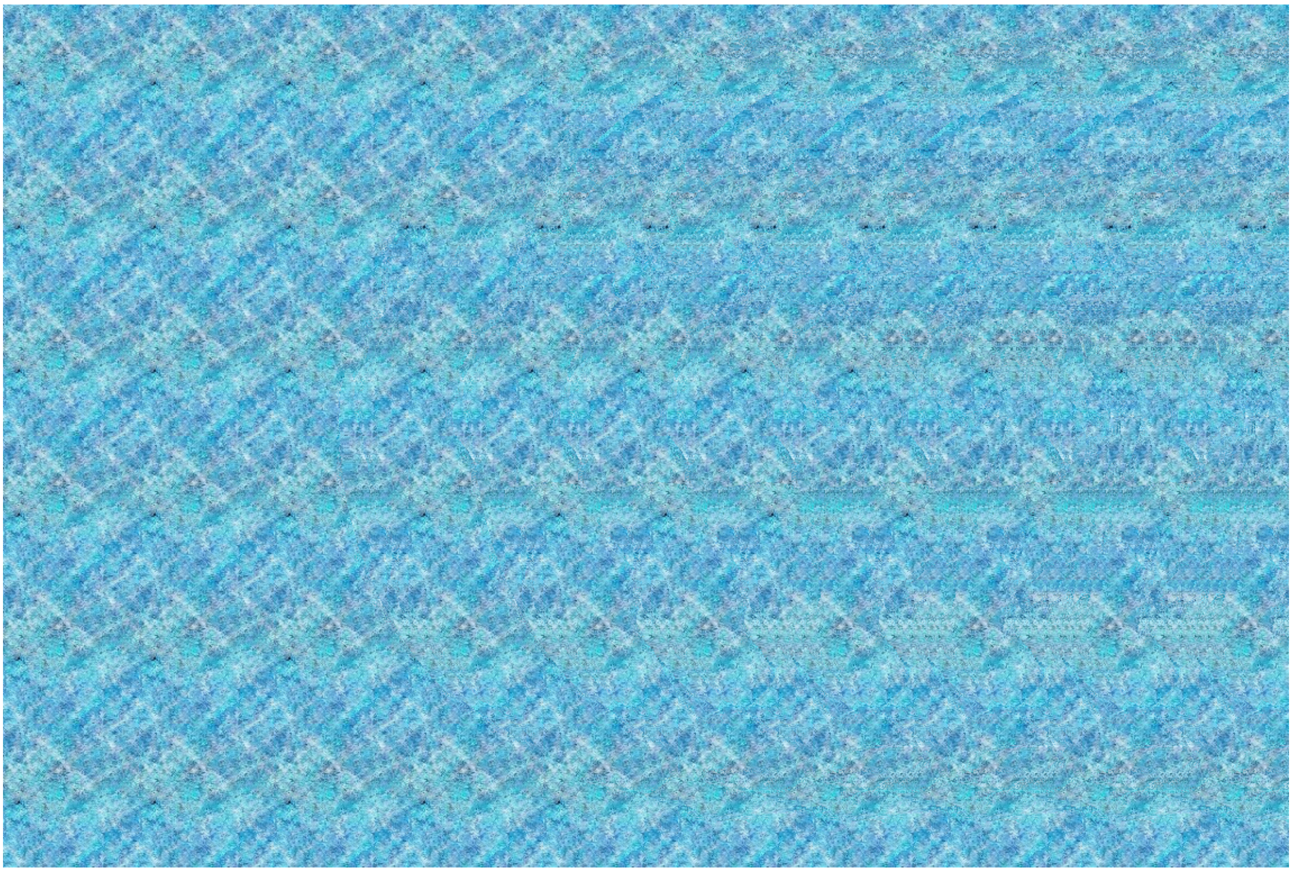


Figure 4.2.4. Stereogram – 'Magic Eye'.

Almost all our senses function by transforming physical interactions with the external world into electrical signals that the brain can process and interpret. These interactions include air vibrations striking the eardrum or photons hitting the retina in the eye. Since the brain operates almost entirely through electrical signals, it relies on this transformation to create our perceptions. Interestingly, these electrical signals are very different from what we ultimately experience. Our perceptions are not raw representations of the sensory data but are transformed into richer, more meaningful experiences by the brain.

For example, light is nothing more than electromagnetic radiation composed of vibrating energy packets called photons. These photons vibrate at different wavelengths, and the retina in our eyes reacts to these variations, giving us the experience of sight and colour. Similarly, sound is the brain's interpretation of air vibrations that hit the eardrum. Variations in the wavelengths of these vibrations result in the differences in pitch that we perceive. This explains why space is completely silent, as without air, there are no vibrations for our ears to detect, and sound cannot exist. As a result, the world outside of our sensory systems and the brain's creative processing is inherently silent and colourless. Much of what we experience, such as colour and sound, is a product of top-down processing. Ironically, the most accurate depiction of reality might be the old black-and-white silent films. This insight highlights how remarkably creative our brain is in constructing a vibrant, colourful, and sound-filled world.

Our minds curate sensory information much like a museum curator organises exhibits to tell a cohesive story. This can be observed in how psychological states influence perception. For example, when we are under stress, our sensory and perceptual systems shift focus to prioritise specific signals. A recent study even found that stress impairs our ability to visually distinguish scenes (Paul et al., 2016). This phenomenon has inspired numerous social media videos showcasing amusing pranks that take advantage of altered perceptions under stress.

In addition to the brain's top-down creative input, it is worth noting that our sensory organs are only sensitive to a narrow range of signals. For instance, Figure 4.2.5 illustrates the tiny fraction of the electromagnetic spectrum that our eyes can detect, which is less than a ten trillionth of all light waves. Similarly, our hearing is limited to a small range of frequencies, from 20 Hz to 20 kHz. Air vibrates at countless frequencies, yet our ears and brains only respond to a narrow slice of this spectrum. This means the world is flooded with information and signals we are ordinarily unaware of.

Fortunately, humans have developed powerful instruments to extend our sensory capabilities, enabling us to detect more of this hidden information. However, these tools come with their own unique challenges and limitations in perception, which we will not delve into here.

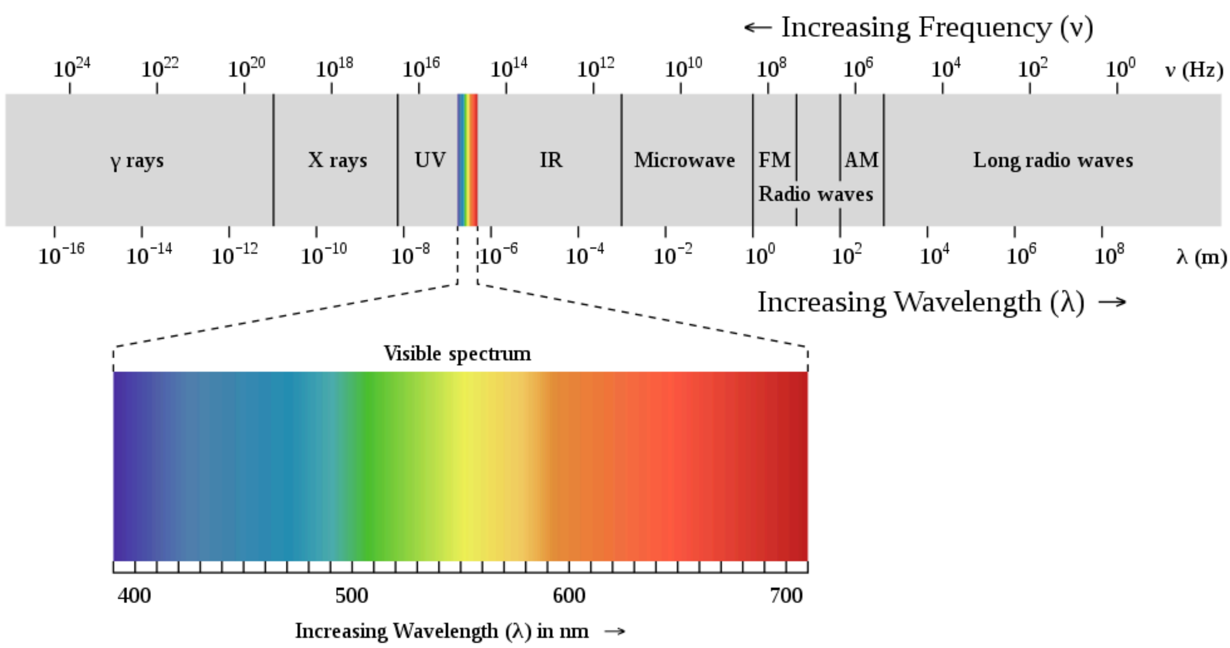


Figure 4.2.5. The Spectrum of Electromagnetic Radiation detected by Our Eyes by EM Spectrum is used under a CC BY-SA 3.0 licence

This discussion highlights key points a critical thinker should consider when understanding the basic processes of sensation and perception. One important fact to keep in mind is that these processes do not work the same way for everyone. People often have different experiences of the exact same thing. A striking example of this occurred in 2015 with the viral social media phenomenon #TheDress. This image revealed significant differences in colour perception among individuals. Some viewers saw the dress as blue and black, while others perceived it as white and gold, and even now, scientists are not entirely sure why. One theory suggests that our eyes adapt to different types of lighting depending on our lifestyle. For instance, people who are night owls might be more accustomed to dim, cool lighting and see the dress as blue and black, while early birds, accustomed to brighter, warmer lighting, may perceive it as white and gold.

This phenomenon illustrates that there are more individual differences in how our perceptual systems function than we might have previously realised. It can be unsettling to recognise that two people with perfectly functioning vision can look at the same object and see entirely different things.

Given these insights, the old saying "Seeing is believing" becomes far less certain. This should prompt us to approach what we perceive with modesty and acknowledge the limitations and fallibility of our sense organs in shaping our beliefs about the world. Recognising the vulnerabilities of our sensory systems is the first step toward becoming a more critical consumer of sensory input. Understanding that we rarely see the whole picture, and often do not even see an accurate one, can help counter overconfidence in our perceptions.

Moreover, being aware of individual differences in sensory and perceptual processes enables us to better understand others. This awareness fosters empathy and helps us appreciate differing beliefs and worldviews. By acknowledging these variations, we become better equipped to navigate contrasting perspectives. Finally, these issues underscore the importance of cultivating the attitudes discussed in the previous chapters, particularly scepticism and open-mindedness when it comes to our sense perceptions.

References

Kindt, M., Van Den Hout, M. (2001). Selective attention and anxiety: A perspective on developmental issues and the causal status. *Journal of Psychopathology and Behavioral Assessment 23*, 193–202. https://doi.org/10.1023/A:1010921405496

Laufer, O., Israeli, D., Paz, R. (2016). Behavioral and neural mechanisms of overgeneralization in anxiety. *Current Biology, 26*(10), 713-722. https://www.cell.com/current-biology/fulltext/S0960-9822(16)00073-7

Paul, M., Lech, R. K., Scheil, J., Dierolf, A. M., Suchan, B., & Wolf, O. T. (2016). Acute stress influences the discrimination of complex scenes and complex faces in young healthy men. *Psychoneuroendocrinology, 66*, 125–129. https://doi.org/10.1016/j.psyneuen.2016.01.007

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

4.3. ATTENTION

By Judith Rafferty, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Our perceptions are significantly shaped by what we choose to focus on (Goldstein, 2019). Attention acts as a filter, determining which parts of our environment we prioritise and how we interpret them. By directing our attention to specific stimuli, we amplify their importance in our minds, while other stimuli fade into the background. This process highlights attention as a critical factor in understanding how people perceive and navigate their environments. However, there are different types of attention, each serving unique purposes. Figure 4.3.1 below illustrates the distinctions between various types of attention.

In this section, we will explore selective attention and divided attention in greater detail. We will also examine specific phenomena related to attention and how they influence perception, ultimately shaping people's experiences.

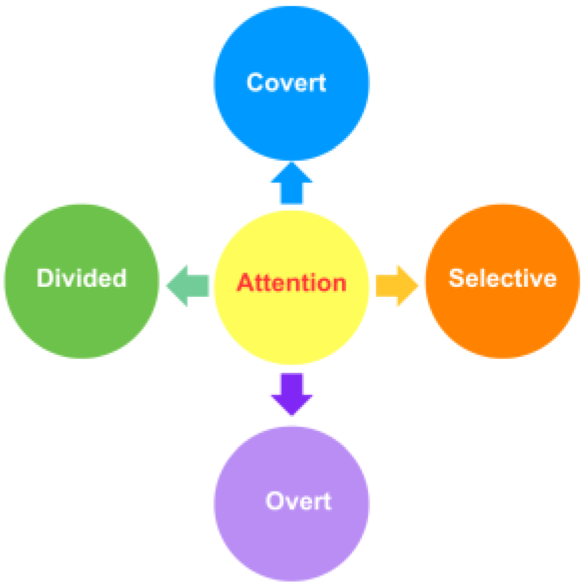


Figure 4.3.1. Different kinds of attention by K. Perry used under a CC BY 4.0 licence

Selective Attention

Selective attention acts like a filter, allowing us to manage the constant flow of information around us.

This process helps us focus on specific inputs for further processing while ignoring others. Heen and Stone (2006) use the example of two individuals, Eric and Fran, who had access to the same pool of information, such as what was said during a work meeting, but each focused on different aspects of that shared information.

The authors describe a phenomenon known as the “user illusion”, which refers to the belief that we perceive everything around us when, in reality, we only take in a small fraction of the available information (p. 344). This selective focus significantly influences how we interpret situations and form conclusions.

In the context of critical thinking, selective attention highlights the importance of being mindful of what we focus on and what we might overlook. The example of Eric and Fran shows how selective attention can shape individual perceptions and influence the narratives we construct about situations. By understanding the filtering nature of attention, we can become more aware of our cognitive biases and work toward a more balanced and comprehensive approach to evaluating information.

Cocktail Party Effect

The cocktail party effect is a fascinating demonstration of selective attention, which is our ability to focus on specific stimuli while filtering out others. Imagine yourself at a lively party: the room buzzes with the chatter of conversations, glasses clinking, music playing in the background, and perhaps the sizzle of food being cooked nearby. Despite all this noise, you can effortlessly concentrate on a conversation with a friend, tuning out the surrounding noise. This ability to selectively focus on one auditory source while ignoring others is what defines the cocktail party effect.

The cocktail party effect hinges on the brain’s ability to process auditory information selectively. While it might seem as though we are ignoring all other sounds, our brains are actually processing these background noises at a lower level. This becomes evident when a stimulus outside our immediate focus grabs our attention. For example, if someone across the room says your name or shouts an alarming word like “Fire!”, your attention will immediately shift to that sound, even though you were not consciously listening for it. This demonstrates the brain’s ability to monitor the environment for meaningful or relevant information while maintaining focus on a primary task.

We rely on selective attention because of our limited cognitive capacity. However, our ability to ignore irrelevant stimuli depends on two factors: the cognitive load of the task we are engaged in and the strength of the task-relevant stimuli. For instance, the Stroop task is a classic example where it becomes challenging to ignore irrelevant stimuli due to the powerful influence of task-relevant information. This interplay of attention and cognitive load reveals how our minds prioritise information in complex environments.

Divided Attention

Divided attention refers to the ability to focus on two or more tasks simultaneously. It occurs when we intentionally split our mental resources to manage multiple activities at once. This ability is essential in many aspects of daily life, allowing us to perform tasks efficiently in environments where competing demands are common.

For example, consider having a conversation while driving. Your attention is divided between processing the road's visual and spatial information, such as monitoring traffic and steering the vehicle, and interpreting and responding to the conversation. Similarly, listening to music while working involves managing the auditory input of the music alongside focusing on the details of your task.

Divided attention is particularly valuable in situations where multitasking is unavoidable, such as managing household chores, engaging in collaborative work, or navigating social interactions while completing errands. However, the degree to which we can successfully divide our attention depends on the complexity of the tasks and the cognitive resources required for each.

While divided attention can seem like a superpower, it comes with significant limitations. Our brain's capacity to process information is finite, meaning that the more tasks we attempt to manage simultaneously, the more likely we are to experience cognitive overload. This can lead to decreased accuracy, slower reaction times, and poorer performance on all tasks involved.

The extent to which divided attention is successful depends on:

1. **Task complexity:** Simpler tasks that are routine or automatic, like folding laundry, are easier to combine with other activities. However, when both tasks require higher levels of focus, such as solving a complex problem while participating in a meeting, performance on one or both tasks will likely suffer.
2. **Cognitive resources:** Some individuals are better at multitasking than others due to differences in cognitive capacity and working memory. However, even those with strong cognitive skills experience limitations when juggling demanding tasks.

It is worth noting that what we often call multitasking is, in many cases, attention switching rather than true divided attention. Instead of processing multiple tasks simultaneously, our brain rapidly alternates focus between tasks. For example, when you check your phone while working on a report, your brain is switching back and forth between the two activities. Each switch requires cognitive effort, and this can create a loss of efficiency known as the switching cost.

Divided attention may be an important skill in today's fast-paced, multitasking world, but it has consequences for our ability to process information effectively. For instance, attempting to multitask during activities that require high levels of focus, such as driving, can increase the likelihood of errors or accidents. Studies have shown that distracted driving, where attention is divided between the road and another task, such as texting, significantly impairs reaction times and decision-making.

Even in less critical scenarios, dividing attention can impact memory retention and the quality of work. For example, listening to a podcast while studying may lead to poor recall of the study material because your brain is splitting its resources between absorbing the audio content and processing the study information.

Inattentional Blindness

Attention is a cornerstone of perception. Without it, we cannot fully process or make sense of the stimuli around us. It allows us to filter the overwhelming amount of sensory information we encounter and focus on what we deem important. However, this focus comes at a cost, as demonstrated by the phenomenon known as inattentional blindness.

Inattentional blindness occurs when we fail to notice a visible and seemingly obvious stimulus because our attention is focused on something else. This phenomenon highlights the limitations of our perceptual systems: while attention helps us concentrate on specific tasks or objects, it also causes us to overlook other, potentially significant details in our environment.

A classic example of inattentional blindness is the experiment conducted by Daniel Simons and Christopher Chabris (1999). If you have not already, take a moment to watch the video below [1:21].

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=1232#oembed-1

In this study, participants were asked to watch a video of people passing a basketball and count how many times players wearing white shirts passed the ball. During the video, a person dressed in a gorilla costume walks through the scene, stops, and even pounds their chest before leaving. Astonishingly, nearly 50% of participants failed to notice the gorilla, even though it was in plain sight.

Inattentional blindness happens because attention is a limited resource. When we direct our focus toward one task, such as counting basketball passes, our brain filters out other information that it deems irrelevant to the task at hand. This filtering mechanism is crucial for managing cognitive overload, but it also means that highly visible stimuli, like the gorilla, can go unnoticed if they fall outside our focus.

This phenomenon illustrates the trade-off in attention: by concentrating on what we believe is important, we lose awareness of other elements in our environment. While this mechanism helps us manage tasks efficiently, it can also cause us to miss critical details.

Inattentional blindness has real-world implications for how we understand perception and attention. It demonstrates that what we perceive is not a complete or objective representation of reality but rather a filtered version based on what we focus on. This insight is critical for understanding human behaviour, decision-making, and even safety.

For example:

- In driving, inattentional blindness can lead to accidents if a driver focuses on a GPS screen or a phone and fails to notice a pedestrian crossing the road.
- In medical settings, inattentional blindness may cause a radiologist to miss an abnormality in a scan if their attention is focused elsewhere.

The Simons and Chabris experiment powerfully illustrates how attention shapes perception. It reveals that perception is not simply a passive process of taking in everything around us but an active one, where attention determines what gets processed and what gets ignored. Understanding inattentional blindness helps us appreciate the limitations of our attention and encourages us to be more mindful of where we direct it. By recognising these limitations, we can strive to broaden our focus when necessary and remain open to noticing details that might otherwise go unnoticed.

References

Goldstein, E. B. (2019). *Cognitive psychology*. Cengage.

Heen, S., & Stone, D. (2006). Perceptions and stories. In A. Kupfer Schneider & C. Honeyman (Eds.), *The negotiator's fieldbook: The desk reference for the experienced negotiator* (pp. 343-350). American Bar Association.

Chapter Attribution

4.4. MEMORY

By Judith Rafferty, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Critical thinkers, educators, decision-makers, and anyone engaging in problem-solving must often handle complex issues and process multiple pieces of information simultaneously. This involves recalling past events, integrating new information, and making sense of it all. These tasks rely on various types of memory, including sensory memory, short-term memory, and long-term memory.

To begin, you might enjoy this video from CrashCourse [9:55], which provides a fun introduction to the topic of memory.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=1237#oembed-1

We will now explore how these different types of memory work.

Sensory Memory

Sensory memory is the initial stage of memory processing, where information from our environment is briefly stored before fading away or being passed along for further processing. This type of memory acts as a buffer, holding sensory input for a very short time, typically just a few seconds or less (Gluck et al., 2020). Sensory memory is crucial for helping us navigate the constant stream of stimuli around us, acting as a bridge between perception and cognition. It allows the brain to decide which information is relevant and worth processing further.

Sensory memory can be divided into several types based on the kind of sensory input being processed.

1. **Iconic memory:** Iconic memory is associated with visual input, allowing us to retain a fleeting image of what we see for less than a second. For instance, when you close your eyes after briefly looking at a bright light, the residual image you "see" is an example of iconic memory.

2. **Echoic memory:** Echoic memory processes auditory input, retains sounds for about three to four seconds. This explains why you can recall the last few words someone said even if you were not fully paying attention.
3. **Haptic memory:** Haptic memory relates to tactile sensations, retaining the immediate sense of touch, such as the texture or pressure you feel when running your hand over a surface.

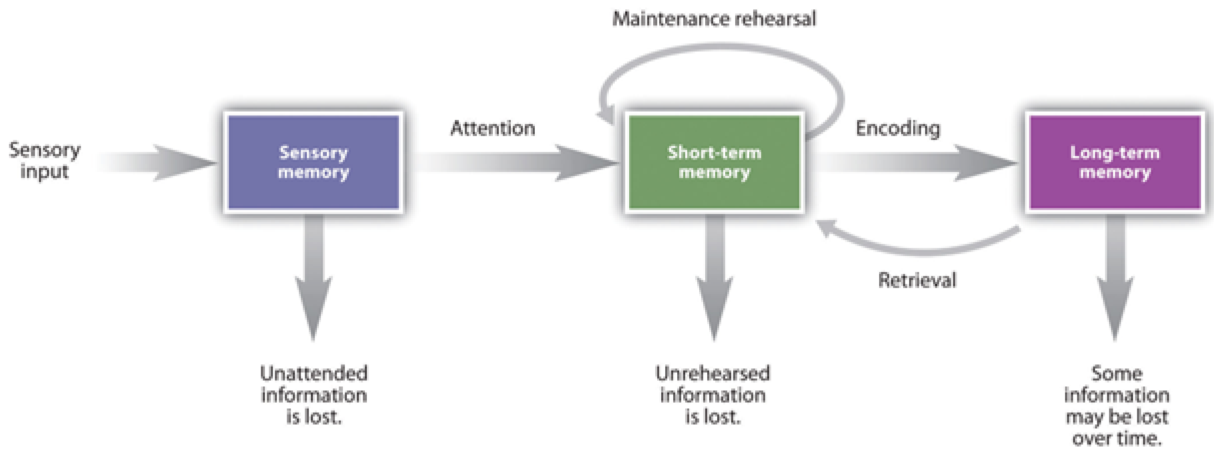


Figure 4.4.1. Memory duration by Jennifer Walinga and Charles Stangor used under a CC BY-NC-SA 4.0 licence

As shown in Figure 4.4.1, attention plays a vital role in sensory memory, determining which sensory inputs move to short-term memory for further processing. Sensory memory holds information for such a brief period that only the stimuli we actively focus on are transferred for further use. This process acts as a filter, helping us manage the vast amount of sensory information we encounter at any given moment. For example, while walking down a busy street, sensory memory registers all the sights, sounds, and smells around you. However, your attention may focus only on the sound of an approaching car or someone waving at you, allowing those specific stimuli to move into short-term memory.

The connection between sensory memory and attention highlights the interconnected nature of cognitive processes. Attention serves as a bridge between perception and memory, determining what sensory input is relevant and worth further processing. For example, perception helps identify and interpret sensory input, such as recognising the sound of a honking car as a potential hazard. Attention then ensures this critical input is passed to memory systems, enabling appropriate action, such as stepping aside to avoid the car.

Short-Term Memory

Short-term memory, also known as working memory, is a critical component of our cognitive system that allows us to temporarily store and manipulate information needed for immediate tasks. It plays an essential role in managing processes like rehearsal, encoding, decision-making, and retrieval strategies, serving as a mental workspace where information is actively held and used. Unlike sensory memory, which lasts only a few seconds, short-term memory can hold information for a slightly longer duration, usually 15 to 30 seconds, before it either fades or is transferred to long-term memory.

Capacity of Short-Term Memory

Research on short-term memory capacity often references the digit-span test, a method in which participants are asked to recall sequences of numbers. This test has revealed that short-term memory typically holds between five to nine items at a time. This capacity aligns with George Miller's famous concept of the "magical number 7", which suggests that the average person can focus on about seven items simultaneously. However, individual differences and specific tasks can influence this capacity. Factors such as fatigue, stress, or distractions may reduce how much information short-term memory can effectively manage at any given time.

Strategies for Retaining Information

Short-term memory employs several strategies to retain information, including coding and chunking:

- Coding is the process of representing information in a specific form, such as auditory or visual.
  - Auditory coding involves recalling sounds, such as a melody, the tone of a loved one's voice, or the rhythm of speech. For instance, when you repeat a phone number aloud to remember it, you are using auditory coding.
  - Visual coding entails recalling images or visual details, like the appearance of a person's face or the arrangement of furniture in a room. This type of coding is particularly useful for tasks requiring spatial awareness or visual recognition.
- Chunking is a method of organising smaller units of information into larger, more meaningful groups. This strategy helps reduce cognitive load and maximise memory capacity.
  - For example, instead of remembering the sequence 1, 9, 7, 9, you might group it as 1979, a year that holds personal or historical significance. Similarly, when trying to memorise a shopping list, grouping items into categories (e.g., fruits, vegetables, and dairy) can make them easier to recall.
  - Another effective chunking technique involves creating a narrative. For instance, if you need to remember a list of random words, constructing a short story that incorporates those words can significantly improve retention.

The Importance of Chunking in Critical Thinking

Chunking is not just a memory aid; it is a powerful tool for critical thinking. By organising complex information into manageable chunks, individuals can better retain and process ideas, making it easier to analyse situations, solve problems, and draw logical conclusions.

For example, educators often use chunking to present lessons in structured segments, allowing students to grasp intricate concepts step by step. Similarly, decision-makers and problem-solvers use chunking to break down complex scenarios into smaller, interconnected parts, enabling them to approach challenges methodically and with clarity. This ability to organise and prioritise information is essential for crafting strategies, forming well-reasoned arguments, and making sound decisions.

Baddeley's Working Memory Model

To better understand how working memory operates, Alan Baddeley developed the working memory model, which divides working memory into several interconnected components. Figure 4.4.2 shows how these components manage different types of information and work together to process and store it. Initially, Baddeley identified three main components, and a fourth, the episodic buffer, was added 25 years later to expand the model:

1. **Phonological loop:** This component handles verbal and auditory information, such as spoken words or sounds.
2. **Visuospatial sketchpad:** This component manages visual and spatial information, such as images or the layout of a room.
3. **Central executive:** Often described as the "manager" of working memory, the central executive oversees information processing. It updates and reorganises memory to balance multiple tasks and switches attention between activities.
4. **Episodic buffer:** Added later, the episodic buffer provides temporary integration of information from the phonological loop, visuospatial sketchpad, and long-term memory. It is controlled by the central executive and serves as a bridge between working memory and long-term memory, facilitating the transfer of information.

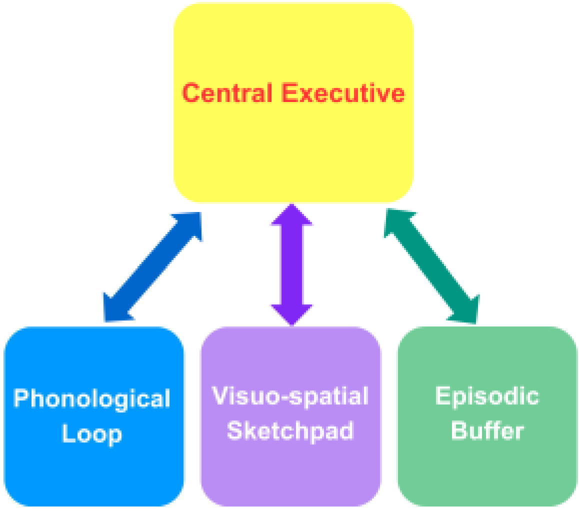


Figure 4.4.2. The four main elements of Baddley's Model of Working Memory by K. Perry used under a CC BY 4.0 licence

To better understand these components, watch this video on working memory by Practical Psychology [7:48]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=1237#oembed-2

As explained in the video, each component of working memory has a limited capacity and operates largely independently. For instance, the visuospatial sketchpad is not affected by the phonological loop, allowing the brain to process visual and auditory information simultaneously without interference.

This model provides practical insights for tasks requiring significant cognitive processing. For example, in situations that involve analysing multiple complex ideas or perspectives, presenting information in different formats can help reduce cognitive load. If verbal explanations overwhelm the phonological loop, visual aids such as diagrams, charts, or illustrations can engage the visuospatial sketchpad instead. For instance, when discussing abstract concepts or exploring interconnected ideas in critical thinking, a teacher might use a whiteboard to visually map out relationships between ideas, helping students process and retain the information more effectively.

The working memory model also sheds light on how information moves into long-term memory. This process involves three key stages:

- **Encoding:** The initial step of memorising information, such as a phone number.
- **Storing:** Maintaining the memory over time, often by rehearsing or repeating the information.
- **Retrieval:** Accessing the stored information when it is needed, such as recalling the phone number.

Long-Term Memory

To wrap up our exploration of memory, let us take a closer look at long-term memory, with a particular focus on the phenomenon of priming. As shown in Figure 4.4.3, long-term memory can be broadly categorised into two types: explicit memory and implicit memory (Goldstein, 2019; Gluck et al., 2020).

Explicit memory, also known as declarative memory, includes two subcategories:

1. **Semantic memory** involves the memory of facts and general knowledge.
2. **Episodic memory** is the memory of personal experiences.

Explicit memory is characterised by awareness. It refers to memories that a person can consciously recall and articulate. For instance, remembering the capital of a country or recalling a family vacation are examples of explicit memory. As Gluck et al. (2020) explain, explicit memory "consists of memory of which a person is aware; you know that you know the information" (p. 280).

On the other hand, implicit memory involves memory that operates without conscious awareness. This type of memory includes skills, habits, and processes that are automatically recalled without intentional effort. For example, riding a bike or typing on a keyboard often draws on implicit memory. According to Gluck et al. (2020), implicit memory is defined as "memory that occurs without the learner's awareness" (p. 280).

Priming is a phenomenon associated with implicit memory and operates unconsciously (Goldstein, 2019). Priming occurs when exposure to one stimulus influences how we respond to a subsequent, related stimulus without our conscious awareness. For example, if you recently read an article about critical thinking, you might be more likely to recognise or recall words like "analysis" or "reasoning" more quickly when encountered later.

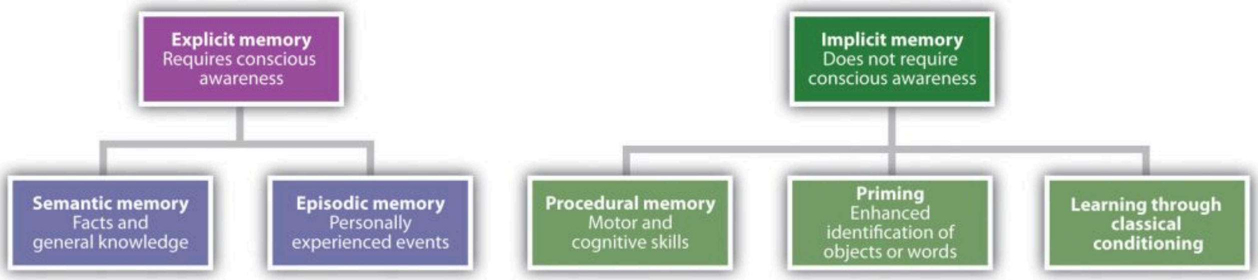


Figure 4.4.3. Types of Memory by Jennifer Walinga and Charles Stangor used under a [CC BY-NC-SA 4.0 licence](#)

Priming

Priming is a psychological phenomenon where exposure to a stimulus influences how we respond to subsequent stimuli and shapes how we perceive and interpret new information. Gluck et al. (2020) define priming as "a phenomenon in which prior exposure to a stimulus can improve the ability to recognise that stimulus later" (p. 88). Similarly, Kassin et al. (2020) explain it as "the tendency for frequently or recently used concepts to come to mind easily and influence the way we interpret new information" (p. 118). Essentially, priming makes certain concepts or ideas feel familiar, even if we are not consciously aware of having encountered them.

For instance, research has demonstrated that when we are subtly exposed to specific words or images, we are more likely to later recognise or choose something related to those stimuli (Gluck et al., 2020; Goldstein, 2019; Kassin et al., 2020). For example, if you are shown words related to "logic" or "analysis" in a subliminal manner, you may be more inclined to approach a problem with a critical thinking mindset.

The Impact of Priming on Social Behaviour

Priming can influence social behaviour by subtly shaping how people act, often without their awareness. This is particularly true when the stimulus is presented subconsciously (Kassin et al., 2020). The impact of priming on behaviour has been demonstrated in various studies, including a notable experiment by Bargh, Chen, and Burrows (1996).

In the first experiment of their study, participants were primed with words associated with either "rudeness" or "politeness". Afterwards, they were placed in a situation where they needed to decide whether to interrupt an experimenter to ask for information. The results showed that participants primed with concepts of rudeness interrupted the experimenter more quickly and frequently than those primed with polite-related stimuli. This demonstrates how subtle cues can influence behaviour in ways consistent with the primed concepts.

In the second experiment, participants were primed with words associated with elderly stereotypes. After the priming, participants who had been exposed to these stereotype-related words walked more slowly down a hallway when leaving the experiment compared to those in the control group. Their behaviour aligned with the traits stereotypically associated with the elderly, showing how priming can subtly alter physical actions based on unconscious associations.

These findings highlight how priming affects not only perceptions but also behaviours in ways that often go unnoticed by the individuals involved. To explore a similar study on the behavioural effects of priming, watch this video by Dcreyethink for further insights [5:12]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=1237#oembed-3

How Priming Can Affect Perception

If you watched the Crash Course video Perceiving is Believing, you may recall an example where viewers were primed to see either a rabbit or a duck based on the framing of a question ("bird or mammal"). This illustrates how priming influences perception, especially when the information presented is ambiguous. In such cases, we are more likely to rely on top-down processing than bottom-up processing to make sense of the stimulus.

Bottom-up processing begins with sensory receptors, which gather raw information from the environment and send signals to the brain. The brain processes these signals and constructs a perception based on the data. However, when stimuli are ambiguous, the brain often relies on top-down processing, where prior knowledge, experiences, and expectations shape how we interpret incoming information. This type of processing, frequently described as concept- or schema-driven, allows us to make sense of ambiguous stimuli by filling in gaps using mental frameworks.

Priming enhances top-down processing because it makes certain concepts or associations more readily available. When we have recently or frequently encountered specific ideas, they come to mind more easily and influence how we interpret new, unclear information. For example, Kendra Cherry, in her article Priming in Psychology, discusses how the Yanny/Laurel viral phenomenon of 2018 demonstrated this effect. The priming effect influenced whether people heard "Yanny" or "Laurel" when confronted with the ambiguous auditory clip.

In the context of visual perception, Feldman Barrett (2017) explains how priming can significantly shape how we interpret others' emotions. She notes that facial expressions are often more ambiguous than we might assume, making them particularly susceptible to the influence of priming. For instance, if we are told a person in a photograph is screaming in anger, we are more likely to perceive anger in their expression, even if this interpretation is inaccurate.

The individual might actually be expressing joy, such as celebrating a significant accomplishment like winning a tennis match. In such a case, the facial expression could reflect a mix of positive emotions. However, when primed to expect anger, our perception narrows, leading us to misinterpret the emotions. Providing contextual information, such as the situation surrounding the facial expression, can help us interpret ambiguous stimuli more accurately.

This example highlights how priming shapes our reliance on prior knowledge to make sense of ambiguous information, emphasising the importance of being mindful of how external cues and context influence our perceptions. Developing awareness of these influences is essential for enhancing critical thinking, as it helps us evaluate situations more objectively and accurately.

References

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244. https://doi.org/10.1037/0022-3514.71.2.230

Feldman Barrett, L. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

Gluck, M. A., Mercado, E., & Myers, C. E. (2020). *Learning and memory: From brain to behaviour* (4th ed.). Macmillan International.

Goldstein, E. B. (2019). *Cognitive psychology*. Cengage.

Kassin, S., Fein, S., Markus, H. R., McBain, K. A., & Williams, L. (2020). *Social psychology* (2nd Australian and New Zealand ed). Cengage Learning.

Chapter Attribution

4.5. BELIEFS

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

A belief is simply the acceptance of something as true. It reflects a personal stance we take toward a claim, where we regard it as accurate or factual. Beliefs are connected to propositions, which are statements that can be either true or false, and a concept we have encountered multiple times in Chapter 2. When someone accepts a proposition as true, they are said to hold a belief in it.

Beliefs can encompass a wide range of ideas. They may involve interpretations of events, evaluations of situations, conclusions drawn from evidence, or predictions about the future. Importantly, beliefs do not exist in isolation. Instead, they tend to form interconnected systems, often referred to as belief systems or working models. These systems consist of clusters of beliefs that support and influence one another, helping us make sense of the world.

Belief models serve as frameworks we rely on to interpret experiences, navigate life, and make decisions. By understanding beliefs and their connections within these systems, we can better appreciate how they shape our perspectives and guide our actions.

Beliefs and Models

How do we navigate and interact with the world? When we wake in the morning and our brains begin processing the flood of electrical and chemical signals from our eyes, ears, and body, how do we make sense of it all? How do we know what to do next? Despite our inability to predict the future, we are not paralysed by indecision or overwhelmed by uncertainty. The answer lies in how our brains manage this complexity: through mental representations or working models of the world. Understanding how these models function helps us recognise both their strengths and vulnerabilities.

What Are Mental Models?

Mental models are like internal maps that we rely on to interpret the world. They help us understand who we are, the resources available to us, and how we should navigate our surroundings. These models are built from past experiences and beliefs, and they allow us to grasp the essential features of the world, anticipate events, and forecast the consequences of our actions. In essence, mental models are tools we use to organise our beliefs, process information, and make decisions.

An important insight is that we do not interact directly with the world; instead, our experiences are filtered through these models. This means we live more "inside our heads" than we might realise. Our understanding of the past, our decisions in the present, and our expectations for the future are all shaped by these models, which act as lenses to filter and shape our perceptions.

Lenses: Focus and Distortion

Like physical lenses, mental models focus our perception but also distort it. This distortion is not inherently bad; it is what makes lenses useful. They simplify the overwhelming complexity of reality, allowing us to focus on specific aspects that are most relevant. However, these distortions also introduce biases and prejudices, which can lead to misunderstandings or flawed judgements. Recognising the benefits and limitations of these lenses is crucial for navigating the world effectively.

What Is a Model?

The term model refers to an abstract representation that highlights key features of something while ignoring less relevant details. For instance, a model plane represents aspects of a real plane, such as its shape, dimensions, and colour, while excluding intricate details like its engine components or the exact number of bolts. Similarly, cognitive models simplify reality, using past experiences and knowledge to create representations of the world that help us make predictions and decisions.

Consider the photograph of a train model in Figure 4.5.1. The image captures important features of the train, but many details are left out. This is the essence of a model: it is not the real thing but a simplified and idealised version designed to highlight what is most important.

Figure 4.5.1. Red and grey train miniature by Darren Bockman, used under an Unsplash licence

One key assumption underlying mental models is that the future will resemble the past. This assumption works well most of the time, but can occasionally lead to errors. The greatest strength of mental models, which is their ability to simplify reality, is also their biggest weakness: they are always incomplete and partial. When we rigidly cling to these models, we risk misunderstanding new information or failing to adapt to changing circumstances.

Flexibility is critical. By revising our models based on new experiences and information, we can refine and improve our understanding of the world. Over time, a well-maintained model becomes increasingly sophisticated and effective at predicting events and guiding our behaviour.

Mental models prioritise practical utility over accuracy. This means they evolve to help us survive and succeed rather than to perfectly represent reality. For instance, certain beliefs, such as superstitions, persist because they serve a purpose, even if they are not entirely accurate. Philosophers have explored this idea extensively, with discussions on "false but useful beliefs" highlighting how certain models, though flawed, can still be beneficial.

By filtering out unnecessary details and noise, models allow us to focus on what is most relevant. For example, a clinical psychologist might use diagnostic criteria as models to identify mental illnesses. While these models are helpful, they do not perfectly capture the complexity of individual cases. Similarly, in the fashion industry, a model is used to idealise how clothing might look, even if it does not perfectly represent reality.

Our mental models pass raw sensory information through multiple filters before any perception takes place. These filters, such as prior knowledge, biases, and cultural influences, shape how we interpret the world. Figure 4.5.2 illustrates an example of these filters.

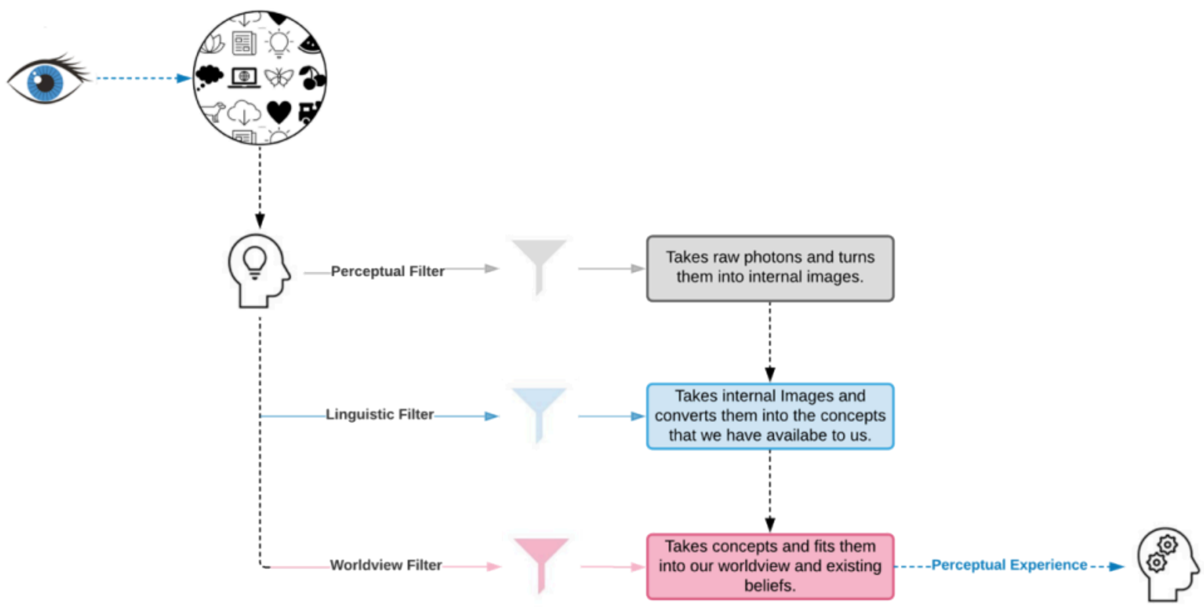


Figure 4.5.2. The series of processing and filtering of visual sensations before a full perception has occurred by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Beliefs and Sensation: A Chicken-and-Egg Problem

The relationship between beliefs and sensation is a dynamic and intricate interaction that defies simple explanations. Many people assume that our beliefs about the world arise solely from raw sensory experiences. However, this perspective oversimplifies the process. Our beliefs not only shape the way we interpret sensory input, but they also act as filters, determining what we sense in the first place. In fact, some beliefs, such as those about abstract concepts like infinity, exist independently of sensory experiences, further illustrating the complexity of this interaction.

How Beliefs Shape Sensation

A classic study by Harvard psychologists Jerome Bruner and Leo Postman (1949) provides a compelling demonstration of how beliefs influence sensory perception. Participants were shown a series of playing cards, some of which were intentionally altered to include anomalies (e.g., a red four of spades, even though spades are always black). When these anomalous cards were briefly displayed, participants often misidentified them as normal cards, confidently reporting what they believed to be "correct". For example, a black six of hearts might be mistaken for a six of spades because participants' existing belief systems filtered the information to fit what they expected to see.

This study illustrates how deeply our beliefs act as lenses or filters, shaping what we perceive and how we make sense of our experiences. Rather than perceiving the world as it truly is, we often see a version of reality that aligns with our pre-existing beliefs.

This interaction creates a feedback loop. Our beliefs influence how we perceive the world, and in turn, our perceptions reinforce our beliefs. Numerous studies confirm that strong belief systems can lead individuals to suppress or dismiss contradictory information. Additionally, our sensory experiences are inherently ambiguous and rely on top-down processing, where the brain uses prior knowledge and expectations to interpret sensory input. This ambiguity makes us susceptible to confirmation bias, a cognitive tendency to favour information that supports our existing beliefs while ignoring or rejecting evidence that contradicts them.

However, ambiguity is not solely a liability. It can also foster creativity and serve as a counterbalance to dogmatism. Embracing ambiguity can open the door to new ways of thinking, encouraging flexibility and curiosity rather than rigid adherence to preconceptions.

Different Belief Systems, Different Worlds

The historian of science Thomas Kuhn, in his seminal work The Structure of Scientific Revolutions

(1962), argued that individuals with different worldviews, or paradigms, effectively inhabit different worlds. Kuhn suggested that people with contrasting belief systems might perceive the same scene in entirely different ways. For example, two individuals with divergent worldviews may interpret the same event, image, or piece of data in fundamentally different ways.

Kuhn went even further to assert that individuals with opposing paradigms might struggle to communicate effectively because they interpret words, concepts, and experiences differently. From their subjective perspectives, it might genuinely seem as though they are living in entirely separate realities.

Although Kuhn's view is extreme, it underscores an essential point: our worldviews significantly shape how we experience and interpret the world. Understanding this can help us navigate conversations and interactions with individuals from different cultures, religions, or political backgrounds, whose worldviews may differ radically from our own.

Our beliefs are not just abstract ideas; they are intimately tied to our sense of self and identity. This is why challenges to our core beliefs can feel deeply unsettling, even threatening. Losing faith in a cherished worldview can be as psychologically devastating as losing religious faith. Beliefs serve as the building blocks of our worldview, and when these are questioned or dismantled, it can feel as though the very foundation of our identity is at stake.

A Circular Dynamic

This relationship between beliefs and sensation is akin to a chicken-and-egg problem: beliefs shape our perceptions, and perceptions, in turn, reinforce our beliefs. As shown in Figure 4.5.3, this dynamic creates a self-reinforcing system in which we interpret the world through the lens of our pre-existing worldview. This process makes us resistant to perceptions or experiences that might challenge our beliefs, a significant liability in critical thinking.

To overcome this tendency, we must consciously work to question and revise our beliefs when faced with new evidence. Being a skilled thinker means adopting an open-minded attitude toward our beliefs and recognising the filters that shape our perceptions. This awareness can help us avoid the pitfalls of confirmation bias, where we seek out and interpret information in ways that confirm our pre-existing beliefs.

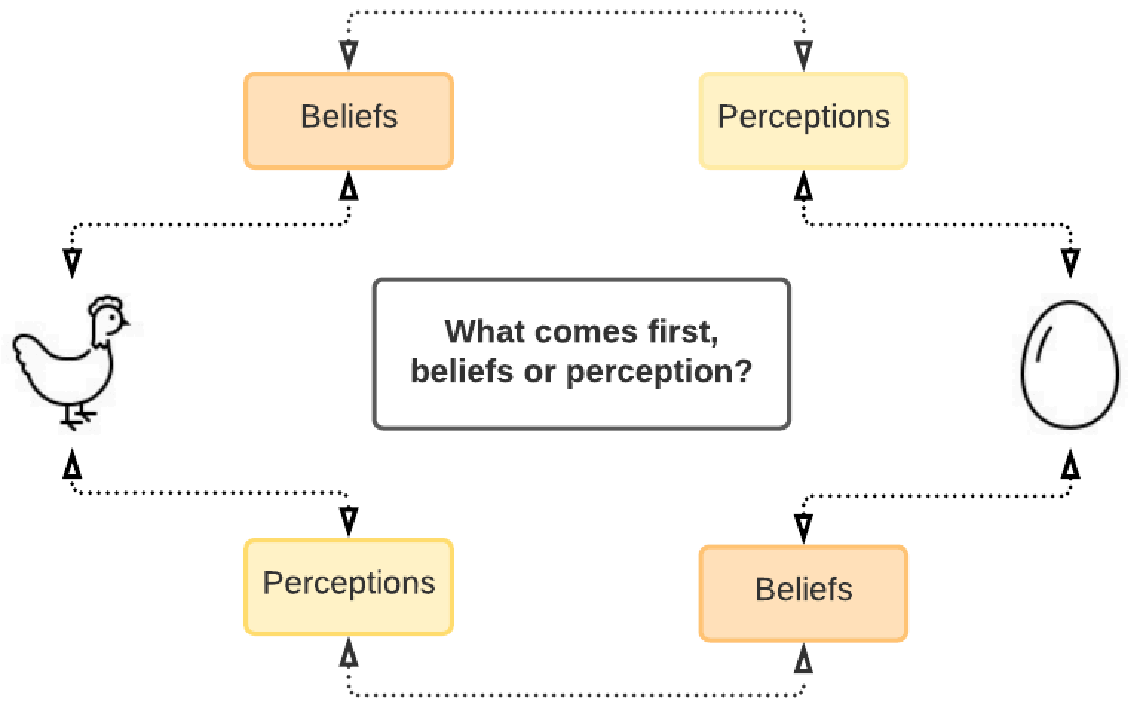


Figure 4.5.3. Beliefs and perception: a chicken/egg conundrum by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Francis Bacon (1561–1626) eloquently articulated the dangers of confirmation bias centuries ago:

> The human understanding, when it has once adopted an opinion, draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects; in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate.

Bacon's insight remains highly relevant today. It reminds us that our natural tendency is to protect and reinforce our beliefs, even in the face of conflicting evidence (Figure 4.5.4). Recognising this bias is the first step toward developing a more flexible and critical approach to our thinking, one that allows us to adapt and grow as we encounter new information and perspectives.

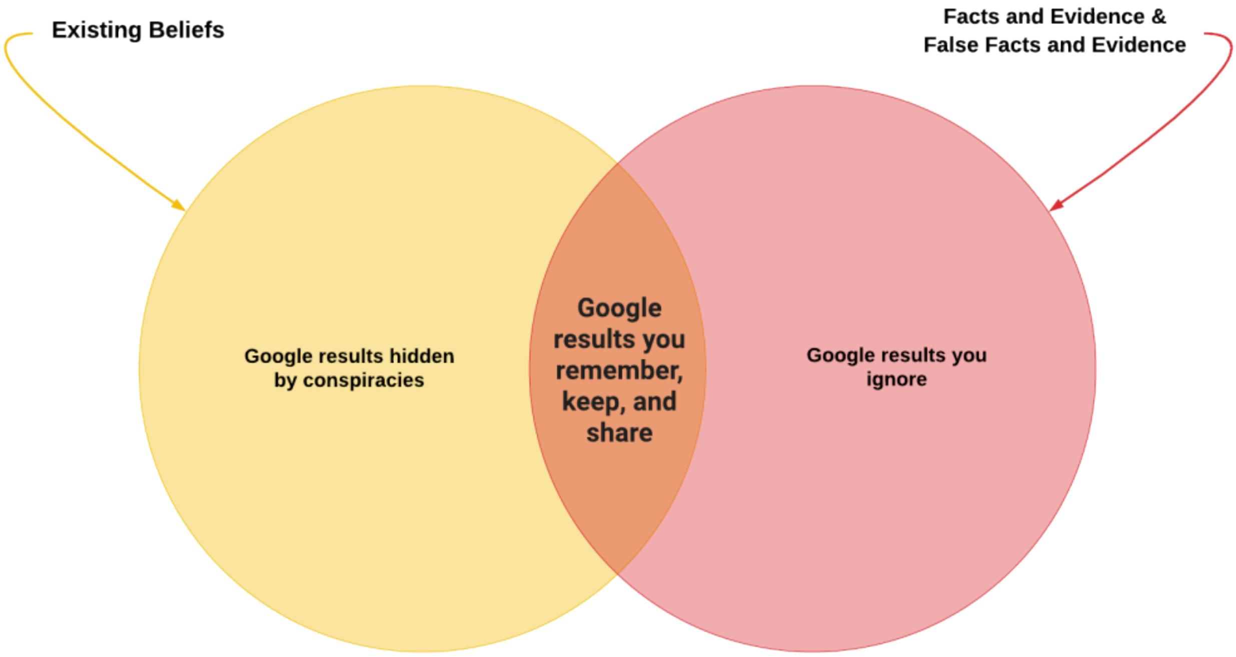


Figure 4.5.4. Our default approaches to interacting with information media by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

Structure of Belief Systems

Beliefs rarely exist in isolation. Instead, they are part of integrated and interdependent systems, which are networks or webs of beliefs that support and reinforce one another. By "integrated", we mean that these beliefs generally fit together in a cohesive way; people tend to resist holding contradictory beliefs. By "interdependent", we mean that the credibility of any single belief often relies on the truth of numerous other beliefs within the system. This interconnected nature of beliefs is described as holistic, an important concept developed by logician Willard Van Orman Quine.

What Is Holism?

In its original sense, holism refers to systems in which individual parts cannot be fully understood in isolation but must be considered as part of the whole. For example, in holistic medicine, the focus is not just on treating the symptoms of a disease but on understanding and addressing the physical, mental, and social factors that contribute to a person's overall health.

When applied to beliefs, holism means that each belief is connected to others in a larger system. This network of interconnected beliefs forms a web (a metaphor introduced by Quine). Beliefs at the core of this web are those most central to our worldview. These are beliefs we are deeply committed to and that are heavily fortified against change. Meanwhile, beliefs on the periphery of the web are less critical and can be more easily modified or discarded without threatening the integrity of the entire system.

The Web of Beliefs

The web structure of beliefs explains why changing someone's deeply held views can be so difficult. Core beliefs are intertwined with many others, forming the foundation of a person's worldview. If a core belief is challenged or disproven, it often triggers a chain reaction that destabilises related beliefs. For this reason, we tend to protect core beliefs, even when presented with evidence that contradicts them.

For example, if someone holds the belief that the Earth is 6,000 years old (a view common among some fundamentalist religious groups), that belief implies a host of others:

- The fossil record must be flawed or misinterpreted.
- Evolution cannot be true.
- Humans and dinosaurs must have coexisted.

These interconnected beliefs create a system where challenging one belief requires addressing the entire network. The person may not even be consciously aware of all the auxiliary beliefs tied to their core worldview until they are confronted with specific challenges.

Quine's web metaphor helps illustrate how we protect our core beliefs. Surrounding the core is a "protective belt" of auxiliary beliefs that can be adjusted or sacrificed to defend the core. When faced with conflicting evidence, people are more likely to abandon peripheral beliefs rather than reconsider those at the centre of their worldview.

For instance, if someone believes the moon landing was faked and is shown satellite images of footprints on the moon, they face a choice:

1. Revise the core belief about the faked moon landing, or
2. Abandon a peripheral belief, such as trust in the accuracy of satellite imagery.

In this scenario, they might choose the second option, dismissing satellite imagery as unreliable, even if they had no reason to doubt it before. By doing so, they protect their core belief, reinforcing their worldview despite the evidence against it. This strategy highlights how we can use an almost infinite number of "moves" to shield deeply held beliefs from revision.

The Emotional Investment in Core Beliefs

Core beliefs are often cherished because they are central to our identity and worldview. Changing these beliefs can feel deeply destabilising or even traumatic. In many cases, these beliefs take on the role of sacred cows, which are ideas we refuse to abandon, even when confronted with overwhelming evidence. Confirmation bias, the tendency to favour information that supports our existing beliefs while dismissing contrary evidence, is one of many psychological tools we use to protect these sacred cows.

Imagine someone who believes the moon landing was faked. This belief might be tied to a broader worldview about government corruption and deceit. When presented with strong evidence, such as images of lunar footprints, they could choose to revise their belief about the moon landing. However, they are more likely to adjust a peripheral belief, such as doubting the reliability of satellite imagery. This allows them to maintain their core belief without disrupting their broader worldview.

Understanding the structure of belief systems highlights why it is so difficult to change deeply held views, whether in ourselves or others. It also shows how people can cling to beliefs despite overwhelming evidence to the contrary, using the web-like nature of their belief systems to defend and reinforce their worldview.

Developing the Right Attitudes Toward Beliefs

With our understanding of how beliefs are formed, structured, and protected, we should now feel motivated to avoid the common pitfalls that our beliefs can lead us into. To do this effectively, we need to adopt a thoughtful and intentional stance toward our beliefs.

Much like the critical thinking dispositions we discussed earlier, this is not an exhaustive list but rather an overview of some helpful approaches to relate to our beliefs in a healthier and more constructive way. These attitudes can guide us in navigating the complexities of belief systems while remaining open to growth and self-reflection.

Modesty

The first essential attitude we need to cultivate is modesty toward our beliefs. After exploring this chapter, several key points should now be clear:

1. We do not know nearly as much as we think we do.
2. Our beliefs are built on sensory and perceptual foundations that are much shakier than we often realise.
3. Our contact with the outside world is indirect, filtered, and augmented.
4. We sometimes unconsciously use devious tactics to protect beliefs we are emotionally attached to.
5. We can be, and likely are, wrong about many of the beliefs we currently hold.

Despite these realities, most of us lack modesty when it comes to our beliefs, and this overconfidence often leads to trouble. Philosopher Willard Van Orman Quine, borrowing a Biblical expression, famously noted that the desire to always be correct is a "pride that goes before destruction", which is a mindset that prevents us from recognising our mistakes and hinders the advancement of knowledge.

Holding Beliefs Lightly

We should approach all our beliefs as provisional, meaning they are rough estimations that have yet to be falsified. Our access to information is always partial and potentially biased, often influenced by someone else's agenda. Therefore, we have no right to absolute certainty about our beliefs.

Unfortunately, much like an insecure lover, many of us feel an overwhelming need for certainty in our beliefs. Without strong critical thinking skills, we find uncertainty intolerable and often cling to bad information, even knowingly, just to avoid the discomfort of not knowing.

Adopting a modest attitude toward our beliefs allows us to:

- Be open to the possibility of being wrong.
- Reduce the impact of confirmation bias, which causes us to seek out and favour information that supports our existing beliefs.
- Foster a more tolerant mindset toward conflicting ideas, as certainty often fuels intolerance.

Certainty can be dangerous, as it often closes us off to alternative perspectives and stifles progress. History shows us that those most certain of their own ideas tend to be the least tolerant of others. By embracing modesty, we create space for growth, learning, and the development of more accurate and inclusive understanding, both individually and collectively.

Falsifiability and Intellectual Courage

Let us revisit two critical points. First, we must always remember that our beliefs are merely our best guesses and are approximations of reality rather than absolute truths. Second, confirming evidence for almost anything is both easy to find and often misleading. In fact, confirming evidence often misses the point entirely. In science, the most valuable evidence is that which emerges from attempts to falsify a proposition and fails to do so. This principle applies not only to scientific theories but also to our everyday beliefs and perceptions.

Confirming evidence, though often comforting, is not the most reliable way to support a belief. This is because it fails to establish whether the belief could withstand serious scrutiny or opposing evidence. Logically speaking, only disconfirming evidence has the proper relationship with propositions, as it directly challenges their validity.

The tendency to rely on confirming evidence is tied to confirmation bias, our natural inclination to seek out information that supports what we already believe while ignoring contradictory evidence. This bias makes it all the more important to emphasise falsifiability, which is the practice of actively seeking evidence that could disprove a belief or claim.

Understanding Falsification

Falsification is a method of approaching evidence by focusing on its ability to contradict ideas or claims. Instead of gathering evidence that supports a proposition, falsification prioritises efforts to disprove it. This concept was popularised by philosopher Karl Popper in his book The Logic of Scientific Discovery (1934), where he argued that the true test of a theory is the ability to attempt to prove it wrong.

Popper viewed confirmation as a flawed approach, riddled with both psychological and logical issues. For instance:

- We are naturally skilled at finding evidence that supports our views, often without realising how selective or biased this process can be.
- Disproving a claim is usually much simpler and more conclusive than trying to confirm it.

Take the claim "All swans are white". To confirm this, we would need to observe millions of swans, and even then, we could not be certain we have seen them all. However, falsifying the claim requires only a single observation of a black swan. The discovery of even one black swan immediately disproves the proposition. This demonstrates the power of falsifying evidence compared to the often inconclusive nature of confirming evidence.

The Role of Scepticism

To think critically, we must approach our beliefs with scepticism. This means not taking confirming evidence at face value and being mindful of the emotional and psychological attachments we often have to certain beliefs. Our tendency to protect cherished beliefs from falsification can cloud our judgement, making it harder to recognise when we are wrong.

If we truly value truth, we need the intellectual courage to actively seek out disconfirming evidence, even if it challenges beliefs we hold dear. This requires vigilance and resilience, as having deeply held views falsified can be unsettling or even devastating. Yet, the pursuit of truth demands this level of boldness and integrity.

As the philosopher Immanuel Kant famously stated, "If the truth shall kill them, let them die." Here, we can interpret "them" as referring to our beliefs, assumptions, or ideas. If a belief cannot withstand scrutiny, it deserves to be abandoned in favour of a more accurate understanding.

The opposite of this courageous stance is intellectual laziness or timidity. Fearing the discomfort of being wrong, we may avoid testing our beliefs or confronting disconfirming evidence. This avoidance undermines our ability to grow and learn, leaving us clinging to false or incomplete understandings of the world.

By contrast, intellectual courage means being bold enough to question and challenge even our most cherished beliefs. This willingness to sacrifice outdated or incorrect ideas can lead to a deeper, more accurate understanding of reality, and, in some cases, might even save your life.

Openness and Emotional Distance

To become effective critical thinkers, we must cultivate openness toward the limitations of our perceptual and belief systems. This means being willing to accept that we could be wrong about what we sense, perceive, or believe. It also involves recognising the biases we use to shield our beliefs from falsification, actively seeking out evidence that challenges our ideas, and remaining open to change when new information arises.

Being open to opposing ideas is not just a hallmark of intellectual maturity, it is also fundamental to critical thinking. When we lack openness, we tend to oversimplify or misrepresent views that differ from our own. For instance, in cultural or political debates, the right often caricatures the left, and vice versa, each side creating comforting but distorted fictions about the other. This resistance to nuance stifles understanding and meaningful dialogue.

Emotional Distance

Emotional distance from our beliefs is equally important. When we tie our sense of identity or self-worth too closely to our beliefs, it becomes much harder to accept that those beliefs could be wrong. This emotional attachment fosters resistance to new information, especially if that information contradicts deeply held ideas.

No one enjoys being wrong or discovering that something they believed is false. However, by practising emotional detachment, we can reduce the discomfort of these moments. Actively seeking to falsify our own beliefs helps us identify and discard incorrect ideas, and cultivating emotional distance makes this process less painful.

By maintaining a healthy detachment, we can approach even our core beliefs with a willingness to question and revise them. This openness not only strengthens our ability to think critically but also ensures that we remain flexible and adaptive in the face of new evidence or perspectives.

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

4.6. SUMMARY

By Marc Chao

Summary

Perception is a dynamic process through which we interpret sensory information to make sense of the world, but it is deeply influenced by individual experiences, biases, and cultural factors. While perception may seem automatic, it involves both bottom-up sensory data and top-down processes like prior knowledge and expectations, leading to significant differences in how people interpret the same events. Attention plays a critical role in shaping perception, as selective attention helps us focus on specific stimuli while divided attention enables multitasking, often at the cost of accuracy. Phenomena such as inattentional blindness illustrate how focusing on one task can cause us to overlook obvious stimuli, underscoring the limitations and fallibility of our perceptual systems. Concepts like the Gestalt principles and the impact of biases, such as naïve realism, further reveal how perception is an active and sometimes misleading process.

The interaction between signal and noise in sensory perception demonstrates how our brains detect, filter, and interpret input amidst irrelevant stimuli. Perception is shaped by both bottom-up (data-driven) and top-down (concept-driven) processing, highlighting the creative and interpretive nature of perception. However, sensory systems have inherent limitations, such as their narrow range of detectable signals and susceptibility to biases. Examples like perceptual illusions and viral phenomena such as #TheDress showcase how individual differences in prior knowledge, expectations, and context can drastically alter perception.

Memory and attention are deeply interconnected, playing critical roles in shaping perception and behaviour. Sensory memory acts as a temporary buffer for environmental stimuli, while short-term memory actively retains and manipulates information for immediate tasks. Strategies like chunking and coding enhance memory efficiency, and long-term memory organises information into explicit and implicit categories. Phenomena like priming demonstrate how past experiences influence current perceptions, while inattentional blindness reveals how

attention filters and organises incoming stimuli. These processes illustrate how biases and context shape our understanding of ambiguous or complex information, reinforcing the importance of cognitive processes in perception.

Beliefs are personal stances toward claims we accept as true, forming mental models that guide our perceptions, decisions, and actions. These models act as internal maps, simplifying the complexity of reality but inherently filtering and distorting information through biases. Beliefs shape our interpretations of sensory input, creating feedback loops that reinforce existing worldviews and may lead to confirmation bias. Core beliefs, central to our identity, are heavily defended, while peripheral beliefs are more easily adjusted. Developing effective critical thinking requires humility, openness, and emotional distance, focusing on falsifiability rather than confirmation to challenge assumptions and foster intellectual growth.

CHAPTER 5: KNOWLEDGE AND SCIENCE

Knowledge has been a central theme in philosophy for centuries, with thinkers from Plato to Kant attempting to define its nature and scope. In this chapter, we embark on an exploration of knowledge, beginning with Plato's foundational definition of knowledge as "justified true belief" (JTB). This classic tripartite theory serves as a framework for understanding the interplay between belief, truth, and justification. Three crucial pillars in determining what qualifies as genuine knowledge.

The chapter navigates through the intricacies of these elements, highlighting their interdependence and their role in distinguishing knowledge from mere belief or error. Readers are guided through key philosophical distinctions, such as the difference between belief and knowledge, the importance of justification, and the challenges posed by subjective interpretations of truth.

Building on this foundation, the chapter expands to explore various categories of knowledge, procedural, acquaintance, and propositional, and the different ways in which justification operates across analytic and synthetic propositions. The chapter also examines the relationship between empirical and theoretical knowledge, drawing on insights from philosophers like Immanuel Kant to emphasise the harmony between sensory experience and rational thought.

Furthermore, the chapter introduces the methods through which humans acquire knowledge, including intuition, authority, rationalism, and empiricism. Each method is critically analysed for its strengths and limitations, showcasing how they contribute to our understanding of the world. The scientific method emerges as a unifying approach, integrating these diverse methods into a systematic, self-correcting process for generating reliable knowledge.

The chapter concludes by distinguishing between science and pseudoscience, emphasising the importance of falsifiability, systematic observation, and public knowledge in maintaining the integrity of scientific inquiry. Readers are encouraged to approach claims with scepticism, evaluate evidence critically, and remain open to uncertainty as a driving force for discovery.

Through this exploration, the chapter not only provides a deep philosophical insight into the nature of knowledge but also equips readers with the tools to evaluate information in an age dominated by competing claims and information overload. Whether you are a student, scholar, or curious thinker, this chapter serves as a guide to understanding how we know what we know, and why it matters.

Learning Objectives

By the end of this chapter, you should be able to:

- **Define knowledge:** Explain the classical definition of knowledge as justified true belief (JTB) and understand the relationship between belief, truth, and justification.
- **Differentiate between belief and knowledge:** Recognise the distinction between personal belief and knowledge supported by evidence and reasoning.
- **Understand the role of justification:** Explain the importance of justification in transforming a belief into knowledge and evaluate different forms of justification.
- **Identify types of knowledge:** Distinguish between procedural, acquaintance, and propositional knowledge and understand their unique characteristics.
- **Differentiate between analytic and synthetic propositions:** Explain the difference between analytic propositions (based on logic and definition) and synthetic propositions (based on observation and experience).
- **Differentiate methods of knowing:** Compare and contrast intuition, authority, rationalism, empiricism, and the scientific method as ways of acquiring knowledge.
- **Identify the goals of science:** Describe the three main goals of scientific research: description, prediction, and explanation.
- **Evaluate scientific research:** Identify the distinguishing features of scientific research, including systematic empiricism, empirical questions, and public knowledge.
- **Understand the difference between science and pseudoscience:** Identify the hallmarks of pseudoscience and explain why distinguishing between science and pseudoscience is crucial.

5.1. KNOWLEDGE

By Michael Ireland, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

The question, "What is knowledge?" has been a central topic in philosophy for centuries. Plato's dialogue, Theaetetus, represents one of the earliest and most influential attempts to define knowledge. While defining knowledge remains a contentious issue among philosophers, one widely discussed definition is Plato's concept of knowledge as "justified true belief", which is often referred to as the JTB definition. According to this view, for someone to claim knowledge, three conditions must be met: the belief must be justified, it must be true, and the individual must believe it. However, as we will see, meeting these conditions is not as straightforward as it might initially seem.

Understanding Belief, Truth, and Justification

Belief is a personal stance or commitment to a proposition about the world. For example, believing that exercise improves mental health is a personal endorsement of that claim. Truth, on the other hand, is more complex because certainty about the truth of many claims is often unattainable. Justification becomes the bridge between belief and truth; it is the reasoning and evidence that support a belief.

The three elements, justification, truth, and belief are closely interwoven. For example, our ability to determine the truth of a belief largely depends on whether we have sufficient justification. Without justification, most people would not consider a belief valid, let alone true.

The distinction between belief and knowledge is crucial. Belief alone is not enough to constitute knowledge. Imagine someone saying, "I don't know how the magician performs that trick, but I believe he uses mirrors." Here, belief represents a personal assumption, while knowledge implies a stronger sense of certainty backed by evidence.

To know something, you must also believe it. For example, if research shows that "a positive attitude toward behaviour increases the likelihood of performing that behaviour", but you refuse to believe this claim, you cannot be said to know it, even though it is true. Similarly, you cannot know something that is not true. If you believe that "a positive attitude has no effect on behaviour", your belief is simply incorrect, not knowledge.

Four Overlapping Categories

The relationship between belief, justification, and truth creates different scenarios:

- **Blind Faith:** Beliefs held without sufficient justification.
- **Denial:** True claims that are rejected despite justification.
- **Errors:** Beliefs that are held but are ultimately false.
- **Knowledge:** Beliefs that are both true and well-justified.

Understanding how these categories overlap helps clarify what separates knowledge from mere belief or error (Figure 5.1.1).

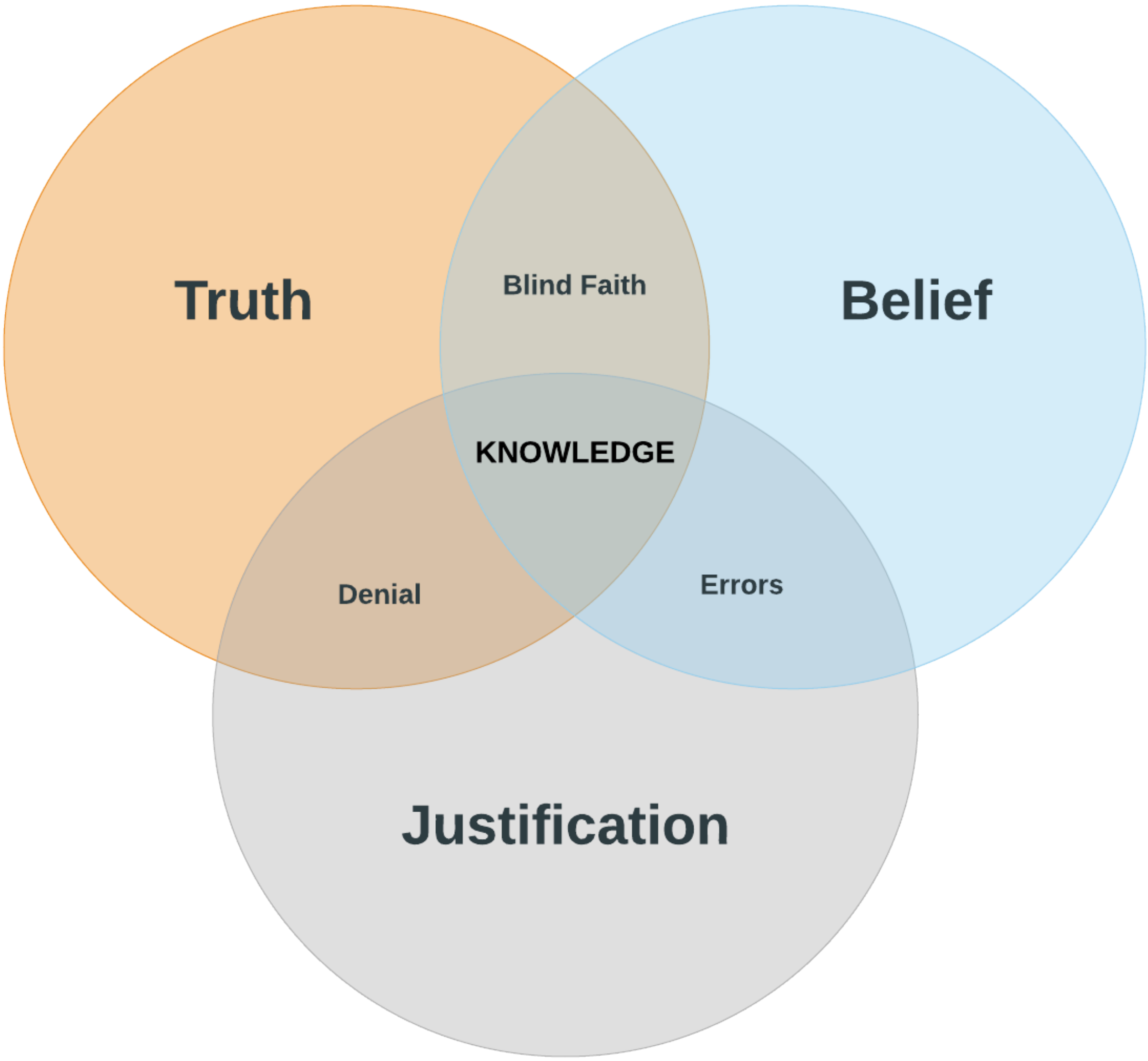


Figure 5.1.1. Locating knowledge at the intersection of justification, truth, and belief by Michael Ireland in Mastering Thinking is used under a CC BY-SA licence

The Challenge of Justification

The most difficult part of the justified true belief model is the justification component. How do we justify our beliefs in a way that transforms them into knowledge? In essence, justified beliefs are conclusions supported by solid reasoning and evidence.

To evaluate whether a belief is justified, we rely on arguments. An argument consists of a conclusion supported by premises (reasons or evidence). To assess justification:

1. **Identify the conclusion:** What is the main point the argument is trying to prove?
2. **Identify the premises:** Ask yourself questions like, "Why is this conclusion believed?" or "What evidence supports this conclusion?"
3. **Evaluate clarity and strength:** If the premises are unclear or unsupported, you are not obligated to accept the conclusion.

Different claims require different types of justifications. For example, a scientific claim might need empirical evidence, while a mathematical claim requires logical proof. Understanding how justification works across different domains is key to distinguishing genuine knowledge from unfounded belief.

Uses of the Term 'Knowledge'

The word knowledge can mean different things depending on how it is used. Broadly speaking, there are three main types: procedural knowledge, acquaintance knowledge, and propositional knowledge.

- **Procedural knowledge** refers to knowing how to do something, like baking a cake or changing a car battery. It involves skills or abilities rather than facts or information.
  Acquaintance knowledge is about familiarity. You might know a person, a city like Brisbane, or your local neighbourhood. It is not about knowing facts but about having personal experience or familiarity with something.
- **Propositional knowledge** is the focus of this section. It refers to knowing that something is true. For example, knowing that water boils at 100°C at sea level or the Earth orbits the Sun. This type of knowledge is expressed through propositions, which are declarative statements about facts or states of affairs.

Propositions and Justification

We first discussed propositions in Chapter 2. To recap, a proposition is a declarative statement that makes a claim about the world, such as "The sky is blue" or "Water is made up of hydrogen and oxygen". Different types of propositions require different types of justification to support them.

Two important categories of propositions are analytic and synthetic propositions:

- **Analytic propositions** are primarily found in mathematics and logic. They are true by definition, such as "All bachelors are unmarried men".
- **Synthetic propositions** are more common in everyday life and scientific studies, including psychology. They are based on observation and experience, such as "Regular exercise improves mental health".

Because psychology and many health sciences rely heavily on synthetic propositions, they depend on empirical evidence for justification.

What Does 'Empirical' Mean?

The term empirical refers to knowledge gained through sense experience, what we can see, hear, smell, taste, or touch. In scientific research, empirical evidence comes from observations, surveys, experiments, or data collection. When you hear the word empirical, think of observation, but remember that it includes all sensory data, not just what can be visually observed.

For example:

- Scientists might collect empirical data by watching animal behaviour in a natural habitat.
- Psychologists might survey people about their emotional experiences.

Empirical justification uses these sensory experiences as evidence to support a proposition. This contrasts with rational justification, which relies on logic, reasoning, and theoretical frameworks rather than direct observation.

The Relationship Between Empirical and Theoretical Knowledge

Knowledge is not just about collecting raw data from our senses, nor is it purely about abstract reasoning. It emerges from the interaction between empirical observation and rational thought. These two forms of justification depend on each other:

- Empirical evidence (sense experience) provides the raw material for knowledge.
- Rational thought (concepts and theories) processes, organises, and interprets that raw material to make it meaningful.

Without rational thought, sensory data would be chaotic and meaningless. Without sensory input, rational thought would be ungrounded and speculative. The philosopher Immanuel Kant captured this balance perfectly when he explained that all knowledge begins with sensory experience, moves through understanding, and ultimately culminates in reason. For Kant, reason is the highest form of knowledge.

He emphasised that sense experience, which he referred to as intuition, and rational thought, which involves concepts and theories, must work together to produce knowledge. In his words, "Thoughts without content are empty; intuitions without concepts are blind. The understanding can intuit nothing, the senses can think nothing. Only through their union can knowledge arise."

In essence, empirical observation provides the raw material for knowledge, while rational reasoning organises and interprets that material into meaningful conclusions. Only when these two elements work together can we form a coherent and reliable understanding of the world.

Facts and Knowledge

Facts are statements about the world that we consider to be true. In essence, a fact is a proposition, or an assertion or claim, that has been sufficiently justified with evidence to make it highly unlikely for a reasonable person to reject it. When a proposition meets this standard of justification, it earns the status of a "fact". For example, if ample evidence supports the claim that water boils at 100°C at sea level, we consider it a fact because it has been repeatedly observed and verified.

However, it is important to note that facts are not merely personal opinions or subjective beliefs. The process of labelling something as a fact follows established procedures for justification, often agreed upon by a community, like scientists or researchers, using shared standards. Even if a fact later turns out to be incorrect, it was still considered factual at the time because it met the accepted criteria for justification. This means there is no such thing as "alternative facts", despite public claims to the contrary.

In short, a fact is a proposition that has been sufficiently supported by evidence to be widely accepted as true. It is similar to how earning a degree in psychology requires meeting certain academic criteria; once those criteria are met, you earn the title of "psychologist".

Not all propositions, however, have achieved this level of justification. A proposition that has not been thoroughly supported yet is often called a hypothesis or conjecture.

A hypothesis is essentially a testable claim about the world. Historically, the term carried a negative connotation. Isaac Newton, for instance, famously said, "I do not feign hypotheses" as a criticism of speculative claims. Today, however, the term hypothesis refers to a tentative explanation for a phenomenon, which can be tested through observation or experimentation.

According to Barnhart (1953), a hypothesis is a single proposition proposed as a potential explanation for an observed event or phenomenon. However, for a hypothesis to be considered scientific, it must be falsifiable. This means it must be possible, in principle, to prove the hypothesis wrong through empirical observation or logical reasoning. If a claim cannot be tested or disproven, it does not qualify as a scientific hypothesis.

References

Barnhart, C. L. (Ed.). (1953). The American college dictionary. Harper.

Chapter Attribution

Content adapted, with editorial changes, from:

Mastering thinking: Reasoning, psychology, and scientific methods (2024) by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA licence.

5.2. METHODS OF KNOWING

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Take a moment to reflect on some of the things you know and how you came to know them. Perhaps you believe making your bed every morning is important because your parents insisted it was a good habit. Maybe you think swans are white because every swan you have ever seen fits that description. Or perhaps you suspect a friend is lying because they seem nervous and avoid eye contact. But how reliable are these sources of knowledge? The ways we acquire knowledge generally fall into five key categories, each with its own strengths and limitations.

Intuition often serves as our internal compass, guiding decisions through gut feelings and instinctive judgements. It feels immediate, persuasive, and deeply personal, but it is also shaped by biases and emotional reasoning, which can lead us astray. Authority, on the other hand, relies on trust in figures we consider knowledgeable, such as our parents, teachers, scientists, or religious leaders. While these sources can offer valuable insights, authority is not infallible and can sometimes perpetuate misinformation or outdated beliefs.

Rationalism takes a more structured approach, emphasising logical reasoning to connect premises and draw conclusions. This method can be incredibly powerful when applied correctly, but its reliability depends entirely on the accuracy of the premises and the logical consistency of the reasoning process. Empiricism, in contrast, emphasises observation and experience as the foundation of knowledge. It encourages us to trust what we can see, hear, and measure, but our senses are not always reliable, and individual experiences can be limited or misleading.

The scientific method stands apart by combining the strengths of these approaches into a systematic process of inquiry. It starts with observation, builds on logical reasoning, relies on controlled experimentation, and demands transparency and reproducibility. While it is not without its limitations, the scientific method has become the gold standard for generating reliable and testable knowledge.

In the sections that follow, we will explore each of these methods in greater detail by examining their advantages, their pitfalls, and how they interact with one another to shape our understanding of the world. By understanding these approaches more deeply, we can better evaluate the information we encounter and make more informed, thoughtful decisions in our everyday lives.

Intuition

Intuition is one of the most common ways we arrive at knowledge, offering a seemingly instinctive and immediate path to understanding situations or making decisions. It often feels like a sudden flash of insight, a gut feeling or an unshakable sense of knowing, without the need for deliberate analysis or evidence. Intuition draws on our past experiences, emotions, and subconscious processing of information to provide us with judgments or solutions that feel both immediate and persuasive. For many, it serves as a guide in moments of uncertainty, helping navigate complex situations where time is limited or the available information is incomplete.

At its core, intuition relies on pattern recognition. Our brains are constantly absorbing and processing information, often without our conscious awareness. Over time, these mental shortcuts, known as heuristics, allow us to quickly identify familiar patterns and make snap decisions based on them. For example, a firefighter might sense imminent danger in a burning building without being able to articulate why, only to realise moments later that the heat or the sound of the fire indicated structural collapse. Similarly, a seasoned chess player might intuitively know the best move in a complicated game scenario, even without systematically analysing every possible outcome.

However, intuition is not without its flaws. While it can be remarkably effective in some contexts, it is also heavily influenced by cognitive and motivational biases. Our gut feelings can be clouded by stereotypes, assumptions, and emotional responses that might not align with objective reality. For instance, if a friend appears distant and avoids eye contact, intuition might suggest they are lying. But in reality, they might simply be exhausted, distracted, or dealing with personal stress unrelated to their interaction with you. These biases can lead to false conclusions when intuition is relied upon without any attempt to verify or cross-check against evidence or reasoning.

Another limitation of intuition is that it thrives on familiarity and prior experience. Intuition often works best in areas where someone has significant expertise or exposure. An experienced doctor might intuitively sense a serious condition based on subtle cues that a less experienced colleague would overlook. However, intuition becomes far less reliable in unfamiliar domains. When dealing with topics or situations outside our expertise, our intuitive judgments are more likely to be driven by guesswork than informed insight.

Despite its imperfections, intuition can still play an important role in decision-making. In some scenarios, especially when time is of the essence, intuitive judgements can outperform slow, deliberate analysis. Overthinking a decision can sometimes lead to "analysis paralysis", where the fear of making the wrong choice prevents any choice at all. In such cases, intuition acts as a valuable shortcut, bypassing unnecessary hesitation and allowing quick, decisive action. For example, an emergency responder might rely on a split-second intuitive judgment to save a life in a chaotic situation where methodical reasoning would take too long.

Furthermore, intuition and analysis are not mutually exclusive; they can work together effectively. Intuition might offer an initial insight or direction, while logical reasoning can be used to validate or refine that insight. In this way, intuition serves as a starting point, a spark that ignites the process of deeper investigation.

Authority

Authority is one of the most common sources of knowledge, rooted in our tendency to trust those we perceive as experts, leaders, or figures of influence. From a young age, we are conditioned to rely on authority figures such as parents, teachers, doctors, religious leaders, scientists, government officials, and media personalities to guide our beliefs and decisions. This reliance on authority is often practical and necessary, as no individual has the time, expertise, or resources to independently verify every piece of information they encounter. For example, most of us accept scientific findings about medicine or climate change because they come from experts with years of education, research, and experience.

Authority, when well-founded, can be an efficient and reliable way to acquire knowledge. Experts often possess specialised training, access to evidence, and analytical tools that allow them to reach conclusions the average person cannot. A pilot, for instance, understands the complexities of aviation in ways a passenger cannot, and a medical doctor can diagnose illnesses based on training and experience that most laypeople lack. Relying on such expertise can save time, prevent errors, and provide access to knowledge that would otherwise be inaccessible.

However, authority is not infallible. History is littered with examples of the dangers of unquestioning obedience to authority. Events like the Salem Witch Trials, where innocent people were executed based on unfounded accusations, or atrocities committed under oppressive regimes, such as Nazi Germany, reveal how authority can be misused or manipulated for harmful purposes. These examples remind us that authority figures are not immune to error, bias, or self-interest.

Even in more benign cases, reliance on authority can sometimes lead us astray. For instance, many of us grew up being told to make our beds every morning because it promotes cleanliness and discipline. However, some studies now suggest that leaving sheets open might actually reduce dust mites by allowing moisture to evaporate. While this example seems trivial compared to historical injustices, it underscores an important point: authority figures, no matter how well-intentioned, can be mistaken, misinformed, or operate based on outdated or anecdotal knowledge.

Moreover, authority is not a monolithic concept; not all authority figures are equally credible, nor are all claims made by experts equally valid. A distinction must be made between legitimate authority, derived from expertise, evidence, and transparent reasoning, and illegitimate authority, which may rely on charisma, fear, or manipulation rather than verifiable knowledge. For example, a climate scientist presenting data from peer-reviewed research holds more authority on global warming than a celebrity expressing personal opinions on the same topic.

Authority can also be undermined by cognitive biases. For example, the halo effect can cause us to overestimate an authority figure's expertise in areas beyond their specific field. A renowned physicist might be a credible source on quantum mechanics, but not on nutrition or political science. Similarly, the bandwagon effect can make people trust authority figures simply because others seem to trust them, creating a false sense of credibility based on popularity rather than evidence.

To make the most of authority as a source of knowledge, it is essential to approach it with a critical mindset. Evaluating an authority figure's credentials, expertise, and track record can provide insight into their reliability. Asking questions such as, "What evidence supports their claims?" or "Do they have any conflicts of interest or biases?" can help uncover potential weaknesses in their arguments. Additionally, credible authorities are usually transparent about their reasoning, provide evidence to back their claims, and are open to scrutiny or peer review.

In an age of information overload, where news headlines, social media influencers, and self-proclaimed experts dominate public discourse, developing a healthy scepticism toward authority is more important than ever. Scepticism, however, does not mean outright dismissal; it means evaluating claims thoughtfully and systematically rather than accepting them blindly. For instance, while it is reasonable to trust a medical professional's advice on vaccines, it is equally reasonable to ask for evidence if their recommendations seem inconsistent with established guidelines.

Rationalism

Rationalism is a foundational approach to acquiring knowledge that emphasises the use of logical reasoning and deductive thinking to arrive at conclusions. Unlike intuition, which relies on feelings and instincts, or authority, which depends on the credibility of others, rationalism seeks to build knowledge systematically by starting with premises, which are statements or assumptions accepted as true, and applying logical rules to derive sound conclusions. This method is often seen in mathematics, philosophy, and theoretical sciences, where reasoning takes precedence over direct observation.

At its core, rationalism operates on the principle that the human mind can discern truths about the world through logical analysis, even without direct sensory experience. For example, if we accept the premise that all swans are white and then encounter a swan, we would logically conclude that it must be white. While this reasoning seems sound, it highlights a significant limitation of rationalism: it is only as reliable as its premises. In reality, not all swans are white; black swans exist in Australia. If the starting premise is false or incomplete, even flawless reasoning will lead to incorrect conclusions. In short, rationalism cannot transcend the limitations of its foundational assumptions.

Another challenge with rationalism lies in the potential for logical errors, especially among individuals who are not formally trained in reasoning or critical thinking. Logical fallacies, which are errors in reasoning that invalidate arguments, can subtly undermine rational conclusions. For example, someone might argue that because two events occurred in succession, the first must have caused the second (post hoc fallacy). Without careful attention to logical structure, even seemingly sound arguments can fall apart upon closer inspection.

Despite these limitations, rationalism remains one of the most effective ways to generate and evaluate knowledge, especially when paired with other methods like empiricism. In science, for instance, rationalism often provides the theoretical foundation upon which empirical experiments are designed. A physicist might use logical reasoning to predict how a particle should behave under certain conditions, and then an experiment is conducted to observe whether the prediction holds true.

One of the strengths of rationalism is its ability to extend knowledge beyond what can be directly observed. Mathematical proofs offer a clear example of this. Take the Pythagorean theorem, which states that in a right-angled triangle, the square of the hypotenuse is equal to the sum of the squares of the other two sides. This truth was derived entirely through logical reasoning, and it holds regardless of whether one ever physically measures the sides of a triangle. Similarly, philosophers have long relied on rationalism to explore abstract concepts such as justice, morality, and free will, building arguments through logical analysis rather than relying on empirical evidence alone.

Rationalism also plays a significant role in everyday reasoning and decision-making. When faced with complex situations, we often use rational thought to weigh options, analyse consequences, and make informed choices. For instance, someone deciding whether to change careers might logically weigh the pros and cons of their current job versus a new opportunity, considering factors such as financial stability, personal fulfilment, and long-term goals. While intuition and emotion may also influence this decision, rational analysis provides a structured framework for evaluating options.

However, it is important to note that rationalism is not immune to cognitive biases. People often unconsciously bend their reasoning to support pre-existing beliefs, a phenomenon known as motivated reasoning. For example, someone who strongly believes in a conspiracy theory might use selective logic to dismiss contradictory evidence while amplifying minor details that align with their views. This demonstrates that even when logical reasoning appears to be at play, it can still be distorted by underlying biases.

To make the most of rationalism as a method of acquiring knowledge, it is crucial to ensure that premises are well-founded, logical rules are consistently applied, and conclusions are critically evaluated. Developing formal reasoning skills through studying logic, argumentation, and critical thinking can greatly enhance one's ability to use rationalism effectively. Additionally, being aware of common logical fallacies and cognitive biases can help individuals spot flaws in both their reasoning and the arguments presented by others.

Empiricism

Empiricism is a foundational approach to acquiring knowledge, emphasising the role of observation, experience, and sensory input as the primary sources of understanding the world. Unlike rationalism, which prioritises logical reasoning, or authority, which depends on the credibility of experts, empiricism insists that knowledge must ultimately be grounded in observable evidence. It is the basis for many scientific discoveries and has significantly shaped modern science, philosophy, and everyday reasoning.

At its core, empiricism operates on the principle that our senses, sight, hearing, touch, taste, and smell, are our primary tools for interacting with and understanding the world around us. For example, if every swan you have ever encountered has been white, you might reasonably conclude that all swans are white based on your observations. Similarly, for centuries, people believed the Earth was flat because, from their limited perspective, the horizon appeared level. These examples highlight both the strengths and limitations of empiricism. On the one hand, it allows us to build conclusions based on real-world evidence; on the other, our observations are inherently limited by our individual experiences, environmental constraints, and the reliability of our senses.

One significant limitation of empiricism is its vulnerability to sensory deception. Optical illusions, for instance, exploit the limitations of our visual perception, leading us to see things that are not actually there. A straight stick might appear bent when partially submerged in water, and the sun seems to rise and set over a stationary Earth, even though we now know the opposite is true. These examples illustrate that while our senses provide valuable information, they are not always reliable on their own.

Additionally, empirical knowledge is constrained by the scope of our personal experiences. No single individual can observe every swan in existence or directly witness every natural phenomenon. As a result, conclusions drawn solely from personal observation are often incomplete and prone to error. This limitation becomes even more pronounced when we consider the role of prior knowledge, expectations, and cognitive biases in shaping how we interpret sensory information. For example, if someone strongly believes in a particular outcome, they may unconsciously focus on observations that confirm their belief while dismissing or overlooking contradictory evidence, a phenomenon known as confirmation bias.

Despite these challenges, empiricism remains one of the most powerful tools for acquiring knowledge, particularly when applied systematically through the scientific method. Scientific empiricism takes the basic principles of observation and experience and elevates them by introducing rigorous methodologies designed to minimise errors, biases, and subjectivity. This approach relies on systematic empiricism, which involves carefully planned and structured observations conducted under controlled conditions. Scientists not only observe phenomena but also design experiments, gather data, and repeat studies to ensure consistency and reliability.

Systematic empiricism also emphasises the importance of falsifiability, which is the idea that for a claim to be scientifically valid, it must be testable and potentially disprovable. For example, the claim "all swans are white" is falsifiable because it can be tested by seeking out non-white swans. When black swans were discovered in Australia, this observation served as empirical evidence that falsified the original claim.

Another key feature of scientific empiricism is the reliance on tools and instruments to extend our sensory capabilities. Telescopes allow us to observe distant galaxies, microscopes reveal cellular structures, and particle accelerators let us study subatomic particles. These tools enable scientists to gather observations far beyond the limits of unaided human senses, providing deeper insights into the natural world.

Empiricism also plays a central role in everyday decision-making. People often rely on past experiences to guide their choices and expectations. For instance, if you burn your hand on a hot stove, you learn through direct experience to be cautious around hot surfaces. Likewise, observing weather patterns might help you decide whether to carry an umbrella. In both cases, knowledge is derived from sensory experience and personal observation, demonstrating empiricism's practical value in daily life.

However, empiricism is most effective when combined with other methods of knowing, such as rationalism and scepticism. While observation can reveal patterns and relationships, rational analysis helps us interpret these patterns and draw meaningful conclusions. For example, observing that the sun rises in the east every morning is an empirical observation, but understanding why this happens requires rational analysis and theoretical reasoning about Earth's rotation.

It is also important to recognise that empirical evidence is not immune to misinterpretation or manipulation. Selective presentation of empirical data, often referred to as cherry-picking, can lead to misleading conclusions. For instance, a marketing campaign might highlight one positive study about a product while ignoring several negative ones. This emphasises the need for critical thinking and transparency in the interpretation and communication of empirical findings.

The Scientific Method

The scientific method stands as one of humanity's most powerful tools for understanding and explaining the world. It integrates the strengths of intuition, authority, rationalism, and empiricism into a structured and systematic approach to knowledge acquisition. While other methods of knowing may serve as starting points for generating ideas or hypotheses, the scientific method goes beyond them by demanding rigorous testing, careful observation, and logical reasoning to ensure that conclusions are well-supported by evidence.

At its core, the scientific method is a cyclical process. It often begins with an observation or question, which is something noticed in the natural world that sparks curiosity or concern. Scientists may rely on intuition to generate initial ideas, turn to authority for background knowledge, or apply rationalism to form logical hypotheses. For instance, a researcher might observe that plants in one area grow taller than those in another and hypothesise that differences in sunlight exposure are responsible. However, rather than stopping at speculation or anecdotal evidence, the scientific method requires a clear and testable hypothesis that can be systematically examined.

Once a hypothesis is formulated, scientists design experiments or studies to test it. This stage involves systematic empiricism, where observations are made in a controlled and repeatable manner to minimise the influence of bias, chance, or extraneous factors. In the plant growth example, the researcher might grow two groups of plants, one in full sunlight and one in partial shade, while controlling for other variables like water, soil quality, and temperature. By isolating the variable of sunlight, scientists can more confidently determine its effect on plant growth.

A defining feature of the scientific method is its emphasis on falsifiability. For a hypothesis to be scientifically valid, it must be possible to prove it false through observation or experimentation. This principle ensures that scientific claims remain open to scrutiny and revision, preventing dogmatic adherence to ideas that cannot be challenged. If the experiment reveals no significant difference in plant growth between the two groups, the hypothesis must be rejected or revised, a process that highlights science's self-correcting nature.

After collecting and analysing data, scientists interpret their findings through logical reasoning. This step often involves statistical analysis to determine whether the results are meaningful or simply due to random chance. Conclusions are then drawn based on the evidence, but even at this stage, they remain provisional. Scientific conclusions are not seen as final truths but as the most reliable explanations given the current evidence. Future research may refine, challenge, or expand upon these findings, which is why replication, or the repeating of studies to verify results, is such a vital aspect of the scientific method.

One of the key strengths of the scientific method lies in its transparency. Scientists are expected to document their methods, data, and reasoning in detail so that others can replicate their experiments and verify their conclusions. Peer review, where other experts in the field critically evaluate a study before publication, adds another layer of scrutiny to ensure the integrity and reliability of scientific findings.

However, despite its strengths, the scientific method is not without limitations. It is often time-consuming and resource-intensive, requiring careful planning, funding, and access to specialised tools or environments. Complex experiments may take years or even decades to complete, which can be a significant barrier when urgent solutions are needed. For example, developing and testing vaccines involves multiple phases of clinical trials to ensure safety and efficacy, a process that cannot be rushed without compromising quality.

Another limitation of the scientific method is its restriction to empirical questions, which are those that can be observed, measured, and tested. Questions about morality, ethics, aesthetics, or subjective experiences often fall outside the scope of scientific inquiry. For instance, science can study how the brain responds to music or why certain patterns are universally considered beautiful, but it cannot definitively answer whether one piece of art is "better" than another.

Additionally, while the scientific method strives to minimise bias, it cannot entirely eliminate human subjectivity. Scientists themselves are influenced by their cultural backgrounds, personal beliefs, and funding sources, which can subtly shape how experiments are designed, interpreted, or reported. This is why transparency, peer review, and replication remain so essential, as they act as safeguards against these biases.

Despite these limitations, the scientific method remains unparalleled in its ability to generate reliable knowledge about the natural world. It has driven countless advancements in medicine, technology, and our understanding of the universe. From discovering the structure of DNA to developing life-saving vaccines and exploring distant planets, the scientific method has consistently proven its value in answering some of humanity's most complex questions.

In practice, the scientific method is not a rigid checklist but a flexible, iterative process. Scientists often revisit earlier stages, refine their hypotheses, and design new experiments in response to unexpected results. This adaptability allows science to evolve, improve, and respond to new challenges in an ever-changing world.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

5.3. UNDERSTANDING SCIENCE

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Many people are surprised to learn that psychology is considered a science. While fields like astronomy, biology, and chemistry are widely accepted as scientific disciplines, psychology's status as a science can seem less obvious. To understand why psychology is a science, it is helpful to first consider what astronomy, biology, and chemistry have in common.

At first glance, these sciences appear fundamentally different from one another. Astronomers study celestial bodies, such as planets, stars, and galaxies. Biologists focus on living organisms, exploring everything from microscopic cells to complex ecosystems. Chemists, on the other hand, investigate the properties, composition, and interactions of matter. Even their tools and methods vary widely. A biologist might use a microscope to study cellular structures, an astronomer might rely on a radio telescope to scan distant galaxies, and a chemist might work with a spectrometer to analyse chemical compounds. The expertise required to operate these tools does not necessarily transfer between fields. For example, a chemist would likely be unfamiliar with the techniques used to track animal populations in the wild, just as a biologist might not know how to interpret radio signals from space.

Yet, despite their differences in subject matter and tools, these disciplines share something far more significant: a commitment to a common approach for acquiring knowledge. Science is not defined by what it studies, but rather by how it studies it. At its core, science is a systematic and evidence-based process of inquiry. It relies on careful observation, the development of testable hypotheses, controlled experimentation, and logical reasoning to draw conclusions. Through this method, scientists aim to reduce bias, ensure reproducibility, and arrive at reliable knowledge about the world.

This same approach applies to psychology. While psychology's subject matter, which includes human behaviour, thoughts, and emotions, might seem less tangible or harder to measure than planets or chemical reactions, psychologists use the same principles of systematic observation, hypothesis testing, and logical analysis to uncover patterns and relationships. For example, a psychologist studying memory might design an experiment to test how distractions affect recall. They would carefully observe participants, collect data, and analyse the results statistically to determine if their hypothesis was supported.

It is also important to recognise that science is not just about gathering facts. It is about understanding the relationships between those facts, identifying patterns, and building theories to explain them. Scientific theories are not random guesses; they are well-substantiated explanations grounded in evidence and refined through repeated testing and observation. In psychology, theories such as Pavlov's classical conditioning or Bandura's social learning theory illustrate how empirical research leads to broader insights about behaviour and mental processes.

However, like all sciences, psychology faces challenges. Human behaviour is influenced by countless variables, including culture, biology, environment, and personal experience, making it more complex to study than chemical reactions in a controlled lab setting. Yet, this complexity does not diminish its scientific nature, it simply means that psychologists must be especially careful in their methods, often relying on large sample sizes, replication of studies, and statistical analysis to ensure their findings are valid and reliable.

Features of Science

The scientific approach is built on three fundamental features: systematic empiricism, empirical questions, and public knowledge. These elements collectively form the foundation of modern scientific inquiry, ensuring that knowledge is built on evidence, focused on answerable questions, and openly shared for verification and improvement.

Systematic Empiricism

Empiricism refers to gaining knowledge through observation, but science refines this process into systematic empiricism. Scientists do not rely on casual observations, gut feelings, or anecdotal evidence. Instead, they carefully design studies, control variables, and document their findings in detail to ensure that their observations are reliable and replicable.

Systematic empiricism involves planning observations in a way that reduces bias and error. Every step, from data collection to analysis, follows a structured process. This systematic approach allows scientists to distinguish genuine patterns from random noise and anecdotal coincidences.

For example, a study investigated whether women talk more than men, not by relying on cultural stereotypes or informal impressions, but by systematically recording conversations from a large, diverse sample of participants (Mehl et al., 2007). Their meticulous approach allowed the researchers to analyse the data objectively and confidently challenge existing assumptions. When their findings contradicted popular stereotypes, they trusted their systematic observations rather than cultural narratives.

Systematic empiricism also incorporates tools and technologies that enhance observation, from telescopes and microscopes to advanced statistical software. These tools extend the range and accuracy of what scientists can measure and observe, allowing them to uncover insights that would otherwise remain hidden.

While systematic empiricism is essential, it does not mean scientists dismiss creativity or intuition. In fact, intuition often inspires hypotheses, and creativity helps design innovative experiments. However, intuition alone is not enough, it must be tested against systematic observations. This commitment ensures that scientific conclusions are grounded in evidence rather than assumptions.

Empirical Questions

Science is focused on empirical questions, which are questions about how the world actually works. These are questions that can be answered through observation and measurement, and they are testable, verifiable, and often falsifiable, meaning they can potentially be proven wrong through evidence.

For instance, the question "Do women talk more than men?" is empirical because it can be answered by systematically observing and comparing speech patterns across genders. Scientists can collect data, analyse patterns, and draw conclusions based on the evidence. Similarly, questions like "Does a specific medication reduce anxiety?" or "How does sleep deprivation affect memory?" are empirical because they can be studied systematically.

However, not all questions are empirical, and science has clear boundaries. Questions about morality, aesthetics, or values, such as "Is it morally wrong to tell a lie?" or "What makes a painting beautiful?", are not testable through scientific observation. While science can inform such discussions (e.g., studying the effects of lying on relationships or analysing brain activity when people view art), it cannot resolve them definitively.

Distinguishing between empirical and non-empirical questions is especially important in psychology, where issues of ethics, values, and subjective experience often intersect with scientific inquiry. Researchers must remain aware of these limits and ensure they do not overstep the boundaries of empirical investigation.

Empirical questions also highlight one of science's core strengths: falsifiability. A scientific claim must be open to being proven wrong through evidence. If a claim cannot, in principle, be falsified, it falls outside the scope of science. This principle keeps scientific inquiry honest, focused, and transparent.

Public Knowledge

The third defining feature of science is its commitment to creating public knowledge. Science is not a private endeavour; it is a collaborative, global effort. Once scientists have asked empirical questions, conducted systematic observations, and drawn conclusions, they share their findings with others, typically through publications in peer-reviewed journals.

These scientific papers provide detailed accounts of the study's rationale, methods, results, and conclusions. They allow other scientists to evaluate the research, replicate the findings, and build upon previous work. Increasingly, researchers are publishing in open-access journals, making their work freely available to the public and removing barriers to accessing scientific knowledge.

Public knowledge serves two critical functions:

1. **Science as a Collaborative Process:** Scientific progress is cumulative, built on the work of countless researchers across time and geography. Every study adds a piece to the larger puzzle of understanding the world. A groundbreaking discovery is rarely the result of one isolated study; rather, it emerges from a long history of shared knowledge and collective effort.
2. **Self-Correction:** Science is inherently self-correcting. Even with the best intentions, individual scientists can make mistakes, overlook biases, or draw incorrect conclusions. By publishing their work, they invite others to review, replicate, and, if necessary, challenge their findings. This process helps identify errors, refine methods, and strengthen the reliability of scientific knowledge over time.

A good example of the self-correcting nature of science is the Many Labs Replication Project. Researchers worldwide collaborated to replicate findings from several influential psychological studies. One of these studies, originally conducted by Schnall and colleagues, suggested that handwashing reduces moral judgements. However, replication attempts using larger samples and identical procedures failed to reproduce the original effect. While this does not conclusively disprove the initial finding, it highlights the importance of replication in verifying scientific claims.

The replication effort demonstrates that science is not about defending individual studies or researchers, it is about uncovering the truth through collective effort. The willingness to re-examine findings, refine methods, and adjust conclusions based on new evidence is what sets science apart from other ways of knowing.

The Interconnected Nature of the Three Features

These three features: systematic empiricism, empirical questions, and public knowledge, are deeply interconnected. Systematic empiricism ensures observations are reliable and unbiased. Empirical questions focus scientific efforts on issues that can be tested and measured. Public knowledge guarantees that findings are shared, evaluated, and refined by the broader scientific community.

Together, these features form the backbone of the scientific method, creating a self-correcting system that drives progress and deepens our understanding of the natural world. Whether studying galaxies, chemical reactions, or human behaviour, scientists rely on these principles to build a body of knowledge that is transparent, collaborative, and grounded in evidence.

In embracing these principles, psychology earns its place among the natural sciences. By adopting systematic empiricism, focusing on empirical questions, and contributing to public knowledge, psychology builds a robust and credible understanding of human thought, behaviour, and experience.

Science Versus Pseudoscience

Pseudoscience refers to a collection of beliefs, theories, or practices that claim to be scientific but fail to adhere to the essential principles of scientific inquiry, such as systematic observation, falsifiability, and peer review. At first glance, pseudoscience can appear strikingly similar to genuine science. It often uses complex, scientific-sounding terminology, references studies or research (whether real or fabricated), and relies heavily on anecdotal evidence to build its credibility. However, beneath the surface, pseudoscience lacks the foundational principles that define genuine scientific inquiry. Understanding the difference between science and pseudoscience is crucial, not only for academic purposes but also for making informed decisions in daily life.

Take the theory of biorhythms as an example. This theory suggests that our physical, intellectual, and emotional abilities follow fixed cycles: 23 days for physical abilities, 33 days for intellectual abilities, and 28 days for emotional abilities, which begin at birth and continue for life. Proponents argue that activities requiring peak performance, like exams or athletic competitions, should be scheduled during the "high points" of these cycles. Books, websites, and even mobile apps have been dedicated to tracking and optimising these cycles, often using terms like sinusoidal wave and bioelectricity to sound more scientific. However, despite the apparent sophistication of these claims, scientific evidence consistently fails to support the existence of biorhythms. Beneath the impressive jargon lies a lack of empirical evidence, systematic observation, or falsifiable hypotheses.

A claim, belief, or activity can be considered pseudoscientific if it presents itself as scientific while failing to meet one or more key features of science. These shortcomings typically manifest in three distinct ways:

1. **Lack of Systematic Empiricism:** Systematic empiricism involves carefully planned, recorded, and analysed observations. In science, researchers rely on these structured observations to test their hypotheses and generate evidence-based conclusions. Pseudoscience, however, often lacks this level of rigour. Claims might rely on anecdotal evidence, personal testimonials, or selective data that support a predetermined conclusion while ignoring contradictory evidence. Even when scientific studies exist, pseudoscientific proponents may misinterpret or cherry-pick findings to suit their narrative.
2. **Absence of Public Knowledge:** Science thrives on transparency. Findings are published in peer-reviewed journals, allowing other scientists to replicate experiments, scrutinise methods, and verify conclusions. Pseudoscience, by contrast, often lacks this openness. Claims may be made without any published research, or findings might be hidden from public scrutiny. Without open evaluation and replication, it is impossible to confirm the validity of the claims.
3. **Failure to Address Empirical Questions:** Science deals with empirical questions that can be tested through observation, experimentation, and measurement. Philosopher Karl Popper emphasised that scientific claims must be falsifiable. This means that there must be a way to prove them wrong if they are indeed false. For example, the claim that "women talk more than men" is falsifiable because

systematic observations could support or refute it. In contrast, many pseudoscientific claims are designed to avoid falsifiability. For instance, believers in extrasensory perception (ESP) often argue that psychic powers fail under controlled observation because the powers are supposedly disrupted by scepticism or scrutiny. This creates an unfalsifiable claim that no possible observation can disprove it, rendering it unscientific.

The Implications of Pseudoscience

Understanding the difference between science and pseudoscience is not just an academic exercise. It has practical, real-world implications that affect decision-making, public health, and the credibility of scientific disciplines.

One important reason for distinguishing between science and pseudoscience is that it clarifies the value of core scientific principles. Systematic observation, falsifiability, and peer review are not abstract concepts but are essential safeguards against misinformation, bias, and faulty reasoning. Studying pseudoscience helps highlight the importance of these principles and fosters a deeper appreciation for evidence-based knowledge.

Another critical consideration is the real-world consequences of pseudoscientific beliefs. These beliefs are not always harmless; they can lead to serious, sometimes life-threatening outcomes. For example, individuals suffering from critical illnesses may reject proven medical treatments in favour of pseudoscientific alternatives such as homeopathy, energy healing, or other unverified therapies. Such choices can delay essential medical care, worsen health outcomes, and, in extreme cases, prove fatal.

Pseudoscience is also particularly relevant to the field of psychology. Many pseudoscientific practices overlap with psychological topics, claiming to explain human behaviour, cognition, and emotion. Examples include astrology, graphology (handwriting analysis), and past-life regression therapy. While these practices often appeal to people searching for answers about themselves, they lack the empirical evidence and methodological rigour required of genuine scientific disciplines. For students and professionals in psychology, the ability to distinguish credible scientific findings from pseudoscientific claims is essential for preserving the field's integrity and ensuring that psychological knowledge remains reliable and evidence-based.

Identifying Pseudoscience

Identifying pseudoscience requires a critical mindset and a solid understanding of scientific reasoning. When evaluating a claim, it is essential to ask whether there is systematic evidence supporting it, if the findings have been published and peer-reviewed, and whether the claim is falsifiable, meaning it can be tested and potentially disproven. Additionally, it is important to consider whether contradictory findings are acknowledged and addressed. If the answers to these questions are unclear or unsatisfactory, there is a strong chance the claim falls into the category of pseudoscience.

Pseudoscience often presents itself in familiar forms, blending cultural beliefs, anecdotal evidence, and scientific-sounding terminology to appear credible. Astrology, for instance, claims that celestial bodies influence human personality and fate, while graphology suggests that handwriting can reveal personality traits. Similarly, practices like energy healing propose that physical and emotional ailments can be cured through the manipulation of unseen energy fields, and magnet therapy asserts that magnets can alleviate pain and improve overall health. Despite their popularity, these fields fail to meet the core principles of scientific inquiry, such as systematic empiricism, falsifiability, and transparency through public knowledge.

For those who want to deepen their understanding of pseudoscience and sharpen their critical thinking skills, several valuable resources are available. The Skeptic's Dictionary offers an extensive reference guide to pseudoscientific beliefs, covering topics like cryptozoology, homeopathy, and pyramidology.

References

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

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

5.4. GOALS OF SCIENCE AND COMMON SENSE

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Human curiosity about the world, and ourselves, has been a driving force behind knowledge and discovery for millennia. From ancient philosophers pondering the nature of the human mind to modern scientists decoding the brain's neural pathways, this fundamental curiosity has shaped our quest for understanding. It is likely this same drive that inspired you to study psychology, a field dedicated to unravelling the complexities of human thought, emotion, and behaviour. Science, as a method for systematically exploring these questions, has proven to be the most effective tool for acquiring accurate, reliable, and verifiable knowledge about these intricate phenomena.

In psychology, scientific research forms the backbone of what we know about human behaviour. From the intricate workings of the brain's cortical regions to the principles that guide learning and memory, almost every insight in the field stems from rigorous scientific inquiry. Research has illuminated everything from how we make decisions under pressure to why we are prone to cognitive biases and what motivates acts of altruism. The knowledge housed in a typical introductory psychology textbook represents the cumulative effort of thousands of researchers over many decades, each contributing a piece to the ever-growing puzzle of human nature. Yet, despite these advances, our understanding of human behaviour remains incomplete. Scientific research in psychology is ongoing, with new discoveries continuously reshaping what we know and revealing how much more there is to learn.

The Three Goals of Science

Scientific research in psychology is driven by three interconnected goals: to describe, predict, and explain phenomena. These goals provide a framework for how psychologists approach their investigations and build knowledge over time.

The first goal, to describe, focuses on systematically observing and recording behaviours, events, or phenomena to create a clear and accurate picture of what is happening. This step is foundational because meaningful insights must start with precise and detailed observations. For example, if researchers want to understand why patients use medical marijuana, they might collect data from licensing centres, survey patients, or analyse patient reports. Findings from such descriptive studies have shown that medical marijuana is most commonly used to treat pain, followed by symptoms of anxiety and depression (Sexton et al., 2016). Descriptive research does not necessarily tell us why or how these relationships exist, but it sets the stage for further investigation.

Building on description, the second goal of science is to predict. Once scientists observe consistent relationships between two variables, they can make informed predictions about future behaviour or events. For instance, if research consistently shows that medical marijuana is primarily used for pain management, scientists can reasonably predict that a patient using medical marijuana is likely dealing with pain-related symptoms. While predictions are rarely 100% accurate, they often provide better-than-random accuracy, especially when the observed relationships are robust and well-documented. Prediction allows scientists to anticipate outcomes, develop interventions, and prepare for potential consequences based on established patterns.

The third and most ambitious goal of science is to explain. Describing and predicting are important, but the ultimate objective is to understand why certain phenomena occur. Explanation involves uncovering the causal mechanisms and processes that drive behaviour or events. For example, in the context of medical marijuana, researchers might ask: How does marijuana alleviate pain? Does it work by reducing inflammation, or does it primarily lower the emotional distress associated with pain without affecting its physical intensity? Answering such questions goes beyond surface-level observations and delves into the deeper mechanisms governing behaviour. Scientific explanations aim to connect observations and predictions into a coherent, causal understanding that can be tested, refined, and expanded upon over time.

Basic versus Applied Research

Scientific research in psychology is often categorised into two broad types: basic research and applied research. While they serve distinct purposes, they are deeply interconnected and often inform one another.

Basic research focuses on building fundamental knowledge about psychological processes without a specific practical application in mind. The goal is to uncover general principles and mechanisms that explain how behaviour, cognition, and emotion operate. For example, studies exploring differences in talkativeness between men and women are not necessarily aimed at solving a specific problem but rather at expanding our understanding of communication patterns and gender dynamics. This kind of research forms the foundation upon which applied research is built.

On the other hand, applied research is driven by practical concerns and aims to address real-world problems directly. Studies investigating the effects of cell phone use on driving behaviour are a good example of applied research. Motivated by safety concerns, such research has influenced public policy and led to legislation aimed at reducing distracted driving accidents. Applied research bridges the gap between scientific knowledge and everyday challenges, translating insights from research into tangible solutions.

While it is convenient to separate research into these two categories, the line between basic and applied research is often blurred. Insights from basic research frequently lead to unexpected practical applications, while applied research often generates new theoretical knowledge. For example, basic research on gender differences in talkativeness might eventually inform communication strategies in marriage counselling, and applied research on cell phone use while driving might reveal new insights about attention and cognitive load.

In essence, both basic and applied research play indispensable roles in advancing our understanding of human behaviour. Basic research builds the theoretical foundation, while applied research ensures that this knowledge is used to address pressing societal issues. Together, they create a cycle of discovery and application that drives progress in the field of psychology.

Science and Common Sense

Many people question whether scientific research in psychology is truly necessary. After all, can we not simply rely on common sense or intuition to understand human behaviour? While it is true that we all possess intuitive beliefs about people's thoughts, emotions, and actions, which psychologists refer to as folk psychology, these beliefs are not always accurate. In fact, scientific research has consistently shown that many widely accepted "common-sense" ideas about behaviour are misleading, incomplete, or outright false.

One classic example is the belief that venting anger, through actions like punching a pillow, screaming into the void, or otherwise "letting it out", can reduce feelings of anger. This idea feels deeply intuitive and emotionally satisfying. However, research reveals the opposite: venting anger in these ways tends to intensify feelings of anger rather than alleviate them (Bushman, 2002). Similarly, many people assume that no one would confess to a crime they did not commit unless they were being physically tortured. Yet, extensive psychological studies have shown that false confessions are alarmingly common and can arise from a range of psychological pressures, including fatigue, fear, coercion, and the desire to escape a stressful interrogation (Kassin & Gudjonsson, 2004).

These examples demonstrate a broader truth: intuition, while often helpful in our day-to-day lives, is not a reliable guide when it comes to understanding complex patterns of human behaviour.

Some Common Myths

Psychologist Scott Lilienfeld and his colleagues explored many such misconceptions in their book, [50 Great Myths of Popular Psychology]. The book highlights numerous common-sense beliefs about human behaviour that have been debunked by scientific research. For example, many people believe that humans only use 10% of their brainpower, that most people experience a midlife crisis in their 40s or 50s, or that students learn best when teaching styles match their preferred "learning style". Other myths include the idea that low self-esteem is a primary cause of psychological problems or that psychiatric admissions and crime rates spike during full moons.

Despite being widely accepted and repeated in popular culture, these beliefs fail to stand up to scientific scrutiny. They persist because they are intuitive, often repeated, and sometimes emotionally comforting. But their persistence highlights the need for rigorous scientific investigation to separate fact from fiction.

Why Are Our Intuitions So Often Wrong?

It is a fair question: If common sense is so fallible, why do we rely on it so heavily? Psychological research points to several key reasons.

Firstly, forming accurate beliefs about human behaviour requires careful observation, precise memory, and detailed analysis. These are abilities that our brains are not naturally equipped to perform on a large scale. For instance, accurately counting and averaging the number of words spoken by men and women in various settings, and drawing valid conclusions from those observations, would be nearly impossible to achieve intuitively.

Instead, we rely on mental shortcuts called heuristics to make sense of the world. These shortcuts are often helpful for making quick decisions in everyday life, but they come with significant limitations. One particularly influential shortcut is confirmation bias, which is the tendency to notice and remember evidence that supports our preexisting beliefs while ignoring or forgetting evidence that contradicts them. For example, if someone believes women talk more than men, they will likely notice and remember talkative women and quiet men while overlooking talkative men and quiet women.

Secondly, many beliefs persist because they are emotionally reassuring. For instance, the idea that calorie-restrictive diets are effective for long-term weight loss persists despite scientific evidence to the contrary (Mann et al., 2007). People often cling to such beliefs because they offer hope, reinforce societal ideals, or provide a sense of personal control.

These cognitive biases and emotional reinforcements create a powerful barrier to changing our minds, even in the face of clear evidence.

The Role of Scepticism

Scientists, including psychologists, are just as vulnerable to cognitive biases as anyone else. What sets them apart is their cultivated attitude of scepticism. Scientific scepticism is not about being cynical, dismissive, or distrusting everything. Instead, it involves a deliberate pause to question assumptions, seek alternative explanations, and, most importantly, demand evidence, especially systematically collected empirical evidence, before drawing conclusions.

For example, imagine reading an article claiming that giving children a weekly allowance teaches them financial responsibility. A sceptical approach would not involve dismissing this claim outright but would instead prompt a series of questions:

- What evidence supports this claim?
- Was the claim derived from systematic research?
- Is the author qualified to make this claim?
- Are there alternative explanations for the observed outcomes?

If the issue is particularly significant, a sceptical approach might also involve digging into the research literature to see whether multiple studies support or contradict the claim.

Tolerance for Uncertainty

Another hallmark of scientific thinking is a tolerance for uncertainty. Scientists understand that evidence is often incomplete and that many questions remain unanswered. For instance, there is currently no definitive scientific evidence showing whether receiving a weekly allowance helps children become financially responsible or encourages materialistic behaviour.

While uncertainty can feel frustrating in everyday decision-making, it is a driving force for scientific discovery. Unanswered questions present opportunities for investigation, experimentation, and innovation. In this sense, uncertainty is not a barrier but an invitation; an open door for scientists and students alike to contribute to our growing body of knowledge.

The Limits of Common Sense

Common sense and intuition have their place in everyday decision-making. They allow us to make quick judgements, navigate social situations, and respond effectively to immediate challenges. However, when it comes to understanding complex psychological phenomena, like the roots of human emotion, the effects of social influence, or the cognitive processes behind decision-making, common sense or intuition often falls short.

Scientific research offers a more systematic, objective, and reliable approach to answering these questions. By combining scepticism, tolerance for uncertainty, and a commitment to empirical evidence, psychology transcends the limitations of intuition and folklore.

In the end, the goal is not to reject common sense entirely but to recognise its limitations. Science does not seek to eliminate intuition; it seeks to refine and validate it through careful observation, analysis, and testing. This evidence-based approach not only advances our understanding of human behaviour but also helps us make better-informed decisions in our personal and professional lives.

References

Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin, 28*(6), 724-731. https://doi.org/10.1177/0146167202289002

Kassin, S. M. (1997). The psychology of confession evidence. I(3), 221–233. https://doi.org/10.1037/0003-066X.52.3.221

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest, 5*(2), 33–67. https://doi.org/10.1111/j.1529-1006.2004.00016.x

Mann, T., Tomiyama, A. J., Westling, E., Lew, A.-M., Samuels, B., & Chatman, J. (2007). Medicare's search for effective obesity treatments: Diets are not the answer. I(3), 220–233. https://doi.org/10.1037/0003-066X.62.3.220

Sexton, M., Cuttler, C., Finnell, J. S., & Laurie K. Mischley, L. K. (2016). A cross-sectional survey of medical cannabis users: Patterns of use and perceived efficacy. *Cannabis and Cannabinoid Research, 1*(1), 131-13. https://doi.org/10.1089/can.2016.0007

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

5.5. SUMMARY

By Marc Chao

Summary

The nature of knowledge has long been a central focus of philosophical inquiry, with Plato's concept of "justified true belief" (JTB) offering a foundational framework. According to this model, knowledge is achieved through the interconnection of belief, truth, and justification. Belief represents a personal commitment to a proposition, truth reflects its alignment with reality, and justification provides the reasoning or evidence to support it. These elements create categories like blind faith (belief without justification) and knowledge (justified true beliefs). The term "knowledge" encompasses procedural, acquaintance, and propositional types, with propositional knowledge being most relevant to logical and empirical investigations. Propositions, declarative statements about the world, can be analytic (true by definition) or synthetic (requiring empirical justification). Empirical knowledge relies on sensory observation, while rational thought organises and interprets these observations to build a coherent understanding. As Immanuel Kant emphasised, rational thought and sensory experience work in tandem to transform raw data into meaningful knowledge. Facts emerge when propositions are sufficiently justified by evidence, distinguishing them from opinions, while hypotheses serve as testable claims that bridge observation and theory.

Knowledge acquisition is underpinned by five key approaches, each with its own strengths and limitations: intuition, authority, rationalism, empiricism, and the scientific method. Intuition, relying on gut feelings and pattern recognition, offers quick judgments but is susceptible to biases and errors. Authority, based on trusting experts, is efficient yet vulnerable to misinformation and misuse. Rationalism emphasises logical reasoning, which is powerful but dependent on the validity of its premises. Empiricism focuses on sensory observation and experience, forming the bedrock of scientific inquiry, though it can be constrained by individual perception and potential misinterpretation. The scientific method unites these approaches into a structured process, prioritising observation, experimentation, falsifiability, and replication to

produce reliable, testable knowledge. While the method has limitations, such as addressing only empirical questions and requiring time-intensive processes, it remains an indispensable tool for advancing understanding and solving complex problems by integrating intuition, reasoning, and evidence.

Psychology, often questioned for its scientific status, shares the essential features of scientific inquiry present in disciplines like astronomy and biology: systematic empiricism, empirical focus, and a commitment to public knowledge. Using rigorous methods, psychology explores human behaviour, thought, and emotion, navigating challenges posed by the variability and complexity of its subject matter. This scientific approach separates psychology from pseudoscience, which lacks systematic evidence, falsifiability, and transparency. Pseudoscience, often cloaked in technical jargon, can have harmful consequences, highlighting the importance of critical thinking and adherence to scientific principles. Identifying and challenging pseudoscientific claims ensures that psychology maintains its credibility and commitment to evidence-based understanding.

At the heart of psychology is the drive to describe, predict, and explain human phenomena, fuelled by the same curiosity that has propelled knowledge and discovery across millennia. Unlike common sense or intuition, which often succumb to cognitive biases and inaccuracies, psychology employs the scientific method to uncover patterns and mechanisms governing human behaviour. It distinguishes between basic research, aimed at building foundational knowledge, and applied research, which addresses real-world problems. This distinction ensures that theoretical insights inform practical solutions and vice versa. Psychology also challenges persistent myths, like venting anger or the notion that humans use only 10% of their brains, using rigorous research to debunk such misconceptions. Through scepticism, empirical evidence, and a tolerance for uncertainty, psychologists refine and expand our understanding of human nature, surpassing the limitations of intuition and common sense to foster a more reliable and nuanced perspective on behaviour.

CHAPTER 6: THE SCIENTIFIC METHOD

Scientific research in psychology serves as a powerful tool for understanding human behaviour, thought, and emotion through systematic observation, experimentation, and analysis. This chapter provides a comprehensive overview of how the scientific process operates within the field of psychology, highlighting the iterative and self-correcting nature of research. At its core, the research cycle is portrayed as a dynamic process, beginning with the formulation of a research question, followed by empirical investigation, data analysis, and the dissemination of findings. Each stage is interconnected, contributing to the ongoing refinement and expansion of psychological knowledge.

The chapter emphasises the importance of grounding research in existing literature, showcasing how prior findings guide new questions and inspire future studies. Real-world examples, such as studies on gender differences in talkativeness and the impact of cell phone use on driving performance, illustrate how research emerges from both theoretical curiosity and pressing societal concerns. These examples also highlight how empirical evidence can challenge widely held stereotypes and inform public policy.

Additionally, the text delves into the creative yet structured process of generating meaningful research questions, underscoring the role of informal observations, practical problems, and gaps in the research literature as sources of inspiration. It introduces essential strategies for conducting thorough literature reviews, including the use of academic databases like PsycINFO, and discusses the importance of distinguishing credible scholarly sources from less reliable ones.

Key distinctions between experimental and non-experimental research designs are explored, along with the trade-offs between internal and external validity in laboratory and field research. Readers are also introduced to the hierarchy of evidence, ranging from anecdotal observations to meta-analyses, to help them critically evaluate the strength and reliability of scientific findings.

The chapter further explains the significance of statistical analysis in interpreting research data, differentiating between descriptive and inferential statistics, and discussing the concepts of statistical significance and replicability. Finally, it underscores the importance of clear and transparent reporting, whether through peer-reviewed journal articles, conference presentations, or public outreach efforts, ensuring that scientific knowledge remains accessible, verifiable, and impactful.

In essence, this chapter serves as a roadmap for understanding the scientific process in psychology, bridging the gap between theory and practice. Whether you are a student embarking on your first research project or a curious reader seeking to understand how psychological knowledge is built and refined, this exploration provides a clear and engaging guide to the scientific underpinnings of one of the most dynamic fields of study.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand the scientific method:** Define the scientific method and explain its key stages, including observation, hypothesis formation, experimentation, analysis, and conclusion.
- **Distinguish between experimental and non-experimental research:** Explain the differences between experimental and non-experimental research and their respective strengths and limitations.
- **Recognise the role of theories and hypotheses:** Explain the relationship between theories, hypotheses, and empirical testing in scientific inquiry.
- **Understand variables and operational definitions:** Define variables, distinguish between quantitative and categorical variables, and explain the importance of operational definitions in research.
- **Assess the validity of scientific findings:** Differentiate between internal and external validity and explain their significance in both laboratory and field research.
- **Analyse statistical data:** Differentiate between descriptive and inferential statistics and understand their role in interpreting research results.

6.1. A MODEL OF SCIENTIFIC RESEARCH IN PSYCHOLOGY

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Scientific research in psychology operates as a dynamic and ongoing cycle, continually building upon itself to refine our understanding of human behaviour and mental processes. As shown in Figure 6.1.1 below, this cycle can be broken down into key stages: formulating a research question, conducting an empirical study, analysing data, drawing conclusions, and sharing the findings through publication. Each step plays a crucial role in advancing psychological knowledge, and the cycle often loops back on itself as new discoveries inspire fresh questions and further investigation.

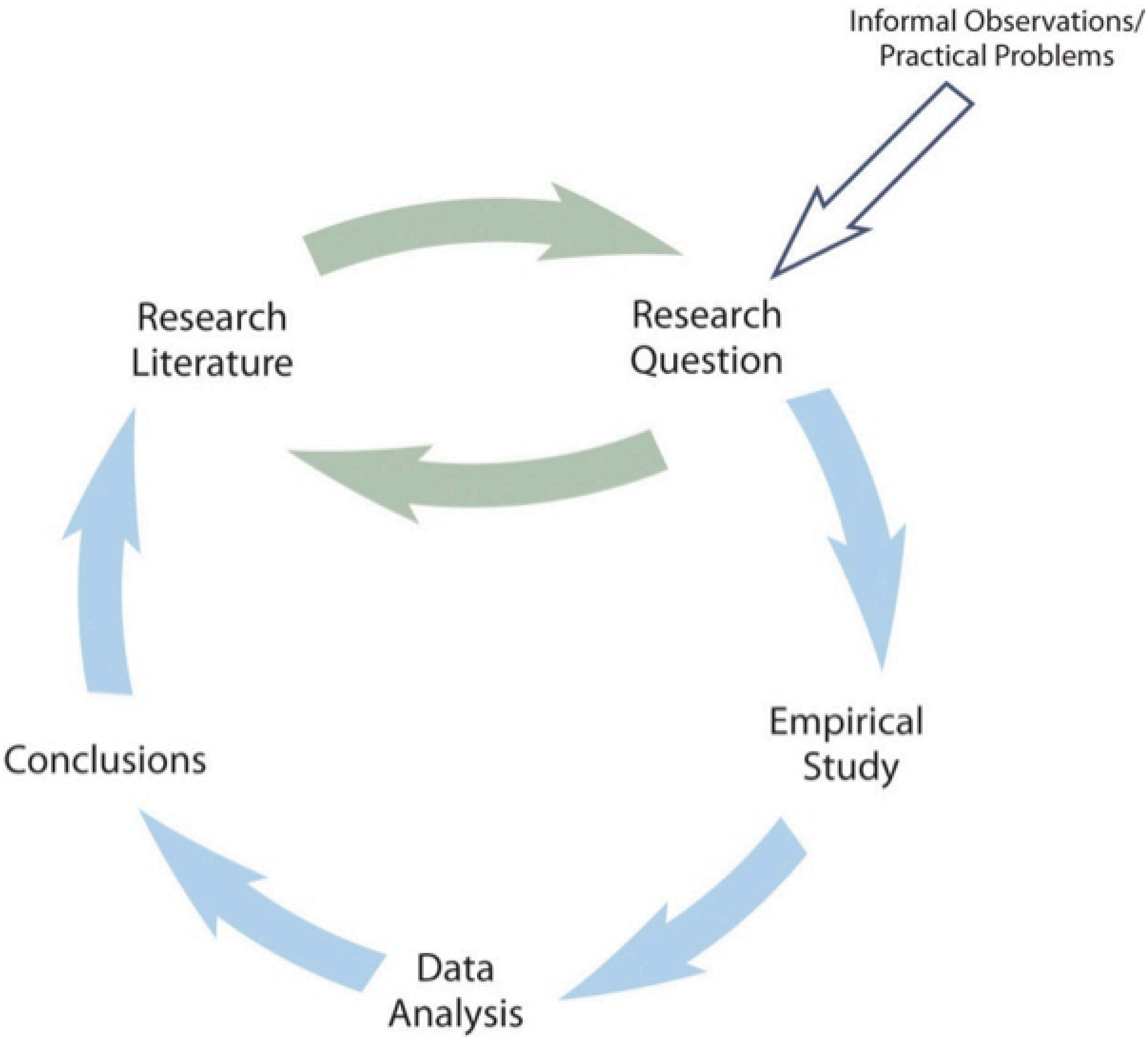


Figure 6.1.1. A simple model of scientific research in psychology by R. S. Jhangiani et al. in [Research Methods in Psychology 4e](#) is used under a [CC BY-NC-SA licence](#)

At the heart of this cycle lies the research literature, which serves as a repository of all published scientific findings within a given field. Researchers rely on this body of work not only to ground their own investigations in established knowledge but also to identify unanswered questions or conflicting results that warrant further study. However, not all research questions originate from the literature. Sometimes, they emerge from casual observations in daily life or from pressing practical problems that demand solutions. Even in these cases, researchers typically begin by consulting the literature to see if their question has already been addressed and to refine their focus based on existing findings.

The Research Cycle in Action

The study conducted by Mehl and his colleagues serves as an excellent example of this research cycle in practice. Their investigation began with a common stereotype: the belief that women are more talkative than men. This stereotype had been perpetuated both informally in everyday conversations and, to some extent, within published claims in the research literature. However, upon closer examination of the literature, Mehl and his team realised that this question had not been adequately addressed by rigorous scientific studies.

With a refined research question in hand, they designed a carefully controlled empirical study. Participants' conversations were systematically recorded (with their consent, of course), quantified, and analysed to determine whether women truly speak more than men. Their results revealed little meaningful difference between the two groups in terms of talkativeness. These findings challenged a widely held stereotype and provided a clear, evidence-based answer to the original question. Importantly, Mehl and his team published their results, contributing to the growing body of research literature on gender differences in communication.

But their study did not mark the end of the story. Like most research, it raised new questions. Could cultural differences affect talkativeness? Are there specific contexts in which one gender might speak more than the other? These follow-up questions offer opportunities for future studies, ensuring that the research cycle continues.

Practical Questions Driving Research

Another example of this cycle comes from a practical, real-world concern: the increasing prevalence of cell phones in the 1990s and their potential impact on driving safety. As cell phone use became widespread, both researchers and the general public began to ask whether talking on a cell phone while driving impaired performance behind the wheel.

Psychologists decided to investigate this question scientifically. Drawing on established research that multitasking tends to reduce efficiency in both tasks, they designed empirical studies comparing driving performance with and without cell phone use. These studies involved both laboratory simulations and real-world driving conditions, allowing researchers to measure drivers' reaction times, hazard detection abilities, and vehicle control.

The findings were clear: cell phone use significantly impaired driving performance. Importantly, the studies went further to reveal nuanced insights. For example, research by Drews et al. (2004) demonstrated that conversations with a passenger were less distracting than cell phone conversations. Passengers in a car are often aware of driving conditions and may naturally adjust their behaviour, pausing their conversation during a challenging traffic situation, for instance. In contrast, someone on the other end of a phone call remains oblivious to the driver's immediate environment, making the conversation more cognitively demanding and distracting.

Each of these studies was published and became part of the growing research literature on distracted driving. These findings did not just advance psychological science; they also had significant real-world applications, informing public policy, raising awareness about distracted driving risks, and influencing laws restricting cell phone use while driving.

For more information about how the brain processes information and what causes driver distraction, watch the following video by the American Psychological Association [3:10]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=911#oembed-1

References

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

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

6.2. FINDING A RESEARCH TOPIC

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Good research starts with a strong research question, but arriving at one can feel like an overwhelming task for novice researchers. Crafting meaningful and testable questions often seems like a mysterious or even magical process, with experienced researchers appearing to pluck ideas out of thin air. However, research on creativity suggests that generating research questions is not an arcane skill but the product of consistent effort, ordinary thinking strategies, and persistence (Weisberg, 1993). In reality, finding a research topic is a creative yet structured process, one that blends curiosity with critical analysis.

Finding Inspiration for Research Ideas

Most research questions begin as broader, more general ideas. These often revolve around behaviours, psychological traits, or phenomena, such as talkativeness, memory, anxiety, or motivation. These initial ideas are then refined and shaped into specific, empirically testable research questions. But where do these initial sparks of curiosity come from?

One of the most common sources of inspiration is informal observation. These are everyday observations of behaviour, whether from personal experiences, interactions with others, or secondhand accounts through books, blogs, news articles, or social media. For example, you might notice that you always seem to pick the slowest-moving line at the grocery store and wonder if everyone feels the same way. Or you might read a story about people rallying to help a family after a fire and start thinking about what motivates people to donate money or resources to strangers. Many famous psychological studies originated from informal observations. Stanley Milgram's groundbreaking research on obedience to authority was sparked by accounts from Nazi war crime trials, where defendants repeatedly claimed they were "just following orders". Milgram turned this observation into a powerful research question: How far will ordinary people go in obeying authority figures, even if it means committing morally questionable acts?

Another significant source of research ideas is practical problems. Practical issues often inspire applied research in areas like health, education, law, and public safety. For example, researchers may ask whether taking handwritten notes improves academic performance compared to typing notes on a laptop. Others might investigate whether psychotherapy is more effective than medication for treating depression, or how cell phone use affects driving ability. These questions stem from real-world concerns and have tangible impacts on people's lives.

However, the most common source of research questions is previous research. Science thrives on collaboration and continuity. Researchers frequently identify gaps, inconsistencies, or intriguing results in existing studies that inspire their own investigations. Experienced researchers often have long lists of potential questions based on their familiarity with the literature. For those just starting out, consulting faculty members or exploring academic journals can provide a wealth of ideas. Simply flipping through a recent issue of a journal like Psychological Science might reveal articles on everything from memory formation and social biases to emotional regulation and second-language acquisition. Focusing on a specific area of interest, such as childhood development or cognitive neuroscience, can further streamline the search.

For more ideas on how to develop a good research topic, watch the following video by KStateLibraries [4:33]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=913#oembed-1

Reviewing the Research Literature

Once a general idea or question has taken shape, the next step is to review the research literature, the body of published scientific studies related to your topic. Reviewing the literature is not just about gathering information; it is about discovering what has already been studied, identifying gaps or inconsistencies, and refining your question to ensure it contributes something meaningful to the field.

A thorough literature review serves several purposes. First, it helps determine whether your question has already been answered. If a question has been thoroughly explored, it might suggest a new direction or angle for investigation. Second, it helps gauge whether the question is interesting and meaningful enough to pursue. Third, it provides insight into methodologies used in similar studies, offering valuable guidance for designing your own research. Finally, it helps clarify how your study will fit into the broader context of existing knowledge.

However, not all sources are equally valuable. The research literature in psychology primarily consists of articles published in professional journals and scholarly books. While popular psychology books, websites, and encyclopaedias may offer insights, they are generally considered unreliable because they are not subjected to rigorous peer review. For example, while Wikipedia can be a useful starting point for general information, its dynamic and anonymous editing process makes it unsuitable as a primary academic source.

Professional Journals: The Core of Research Literature

Professional journals are central to the research literature in psychology. They publish original empirical studies and review articles, serving as a platform for sharing findings, theories, and analyses with the academic community.

Most journal articles fall into one of two categories: empirical research reports and review articles. Empirical research reports describe new studies conducted by the authors, presenting their research question, methodology, results, and conclusions. Review articles, on the other hand, synthesise findings from multiple studies to provide a broader understanding of a topic. Some review articles propose new theories (theoretical articles), while others statistically analyse results from multiple studies (meta-analyses).

What sets professional journals apart is their peer review process. When researchers submit an article to a journal, it undergoes double-blind peer review. This means that the identities of both the authors and reviewers are concealed, ensuring that feedback remains unbiased. Reviewers, who are experts in the field, critically evaluate the study's methods, results, and interpretations, providing constructive feedback and recommendations to the journal editor. The editor then decides whether to accept, revise, or reject the manuscript based on these reviews.

In recent years, open-access journals have gained popularity. These journals make published articles freely available to anyone, removing barriers to accessing scientific knowledge. Some open-access journals also embrace open peer review, where reviewer identities are disclosed alongside published articles to promote transparency and accountability.

Scholarly Books

Scholarly books play a vital role in the dissemination of scientific knowledge in psychology and related fields. These books are primarily written by researchers and practitioners for an audience of fellow researchers, practitioners, and advanced students. Unlike popular psychology books, which are often aimed at a general audience and may lack scientific rigour, scholarly books undergo a more rigorous peer-review process to ensure accuracy, relevance, and quality.

Scholarly books generally fall into two main categories: monographs and edited volumes. A monograph is typically authored by a single researcher or a small group of authors. It provides a deep and coherent exploration of a specific topic, much like an extended review article. Monographs are often structured to build a clear narrative or argument, drawing on extensive research and analysis.

In contrast, edited volumes are collaborative works where an editor, or a small team of editors, curates chapters written by multiple contributors. Each chapter usually explores a different aspect of the overarching topic. While edited volumes can offer a comprehensive overview of a subject, the perspectives presented in each chapter may vary, and sometimes contributors even openly disagree with one another. This diversity of viewpoints can provide valuable insight into ongoing debates and areas of uncertainty within a field.

Both monographs and edited volumes contribute significantly to the research literature, serving as key resources for deep dives into specialised topics, theoretical advancements, or emerging trends in psychology.

Literature Search Strategies

Research in psychology relies heavily on navigating the extensive body of scholarly literature. To effectively find, evaluate, and utilise this literature, researchers use a range of search strategies and specialised tools.

Using PsycINFO and Other Databases

One of the most powerful tools for accessing research literature is PsycINFO, an electronic database maintained by the American Psychological Association (APA). PsycINFO is unparalleled in its comprehensiveness, covering thousands of journals, books, and chapters, with records dating back over a century. For most psychologists, PsycINFO essentially serves as a gateway to the research literature in the field.

Other useful databases include Academic Search Premier, JSTOR, and ProQuest, which cover a broad range of academic disciplines. Specialised databases like ERIC focus on education, while PubMed caters to medicine and health-related fields. Most university libraries provide access to these databases, making them essential tools for any literature search.

Each entry in PsycINFO includes essential details such as publication information, abstracts summarising the research, and lists of cited references. Importantly, entries are tagged with keywords and index terms that help categorise content systematically. For example, all research on sex differences is indexed under "Human Sex Differences", and studies on note-taking strategies are indexed under "Learning Strategies". If you are unsure which terms to use, PsycINFO includes a thesaurus that suggests standardised search terms.

Effective searching often requires refining your search terms. For example, a search for the term "memory" will yield millions of results, making it nearly impossible to sift through them. By consulting the thesaurus and narrowing the term to "early memories" and then combining it with "Human Sex Differences", you can focus your search on highly relevant studies.

Many PsycINFO platforms allow users to save, print, or email search results. Some even provide direct links to full-text versions of articles through databases like PsycARTICLES. If full-text access is not available, you may need to check your library's holdings or request materials through interlibrary loan services. Do not hesitate to ask a librarian for help, they are trained experts in navigating these databases.

To learn how to find specific journal articles using filters in APA PsycInfo, watch the following video by the American Psychological Association [2:26]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=913#oembed-2

Using Additional Search Techniques

While databases like PsycINFO are invaluable, they are not the only tools for conducting a literature search. Additional strategies can enhance your search process. If you find a particularly relevant article or book chapter, do not stop there. Explore its reference list for additional sources cited by the authors. These references often lead to other foundational or complementary studies.

Alternatively, if you have a classic or influential article on your topic, you can use PsycINFO or Google Scholar to see which newer studies have cited it. This technique often reveals the latest developments and discussions building on that foundational work. Google Scholar is another powerful tool. While it includes both scholarly and non-scholarly sources, it can quickly identify academic articles, open-access papers, and researcher profiles.

You might also perform a general internet search, which can sometimes lead to preprint versions of papers, articles hosted on researchers' personal websites, or additional resources not captured by traditional academic databases.

Finally, talking to experts in your field, such as instructors, faculty members, or experienced researchers, can provide invaluable guidance. These individuals often have deep familiarity with the research landscape and can recommend must-read articles, books, or emerging topics worth exploring.

To learn how to use the advanced features on Google Scholar, watch the following video by CLIP [7:02]:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://jcu.pressbooks.pub/critical-thinking-psychology/?p=913#oembed-3

What to Look for in a Literature Search

Not every source you find will be equally valuable. A good literature review focuses on sources that serve four primary purposes: refining your research question, identifying effective research methods, situating your research within existing knowledge, and supporting your argument for why your question is interesting.

When evaluating sources, recentness is often a crucial factor. In fast-moving fields, research published within the past year or two may be most relevant. In more established areas, studies from the past decade may still hold significant value. However, classic studies, which are frequently cited and foundational to a topic, should not be overlooked, even if they are older.

Review articles are particularly valuable at the start of your literature search. These articles provide comprehensive overviews of a topic, summarising key findings, highlighting trends, and identifying areas of debate or uncertainty.

Equally important are empirical research reports that address your specific question or similar ones. These studies often offer concrete examples of methodologies and measurement techniques you can adapt for your own research.

Lastly, look for contextual information that helps explain the significance of your research question. For instance, if your topic is the effect of cell phone use on driving performance, finding statistics on traffic accidents caused by distracted driving can help emphasise the importance of your study.

How Many Sources Are Enough?

The number of sources needed for a literature review varies widely depending on the topic, research goals, and scope of the project. Professional journal articles often cite an average of around 50 sources (Adair & Vohra, 2003). For student projects, the required number may be significantly lower, but the principles for selecting high-quality sources remain the same.

The goal is not to accumulate the largest number of citations but to ensure that the sources you include are relevant, reliable, and valuable for shaping your research question, designing your methodology, and framing your conclusions.

References

Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist, 58*(1), 15–23. https://doi.org/10.1037/0003-066X.58.1.15

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371–378. https://doi.org/10.1037/h0040525

Weisberg, R. W. (1993). *Creativity: Beyond the myth of genius*. W. H. Freeman.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology ,(4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

6.3. THE HIERARCHY OF EVIDENCE

By Marc Chao and Muhamad Alif Bin Ibrahim

In scientific research, not all evidence carries the same weight. While every source of information may contribute something to our understanding of a topic, the reliability, validity, and overall usefulness of these sources vary significantly. At the foundation of this hierarchy (Figure 6.3.1) lie anecdotal observations and personal experiences, while at the peak stand meta-analyses and systematic reviews, which are sources that synthesise vast amounts of data to arrive at highly robust conclusions. Understanding this hierarchy is essential for anyone engaging with scientific research, as it helps prioritise the most credible and reliable sources when building arguments, making decisions, or advancing knowledge.

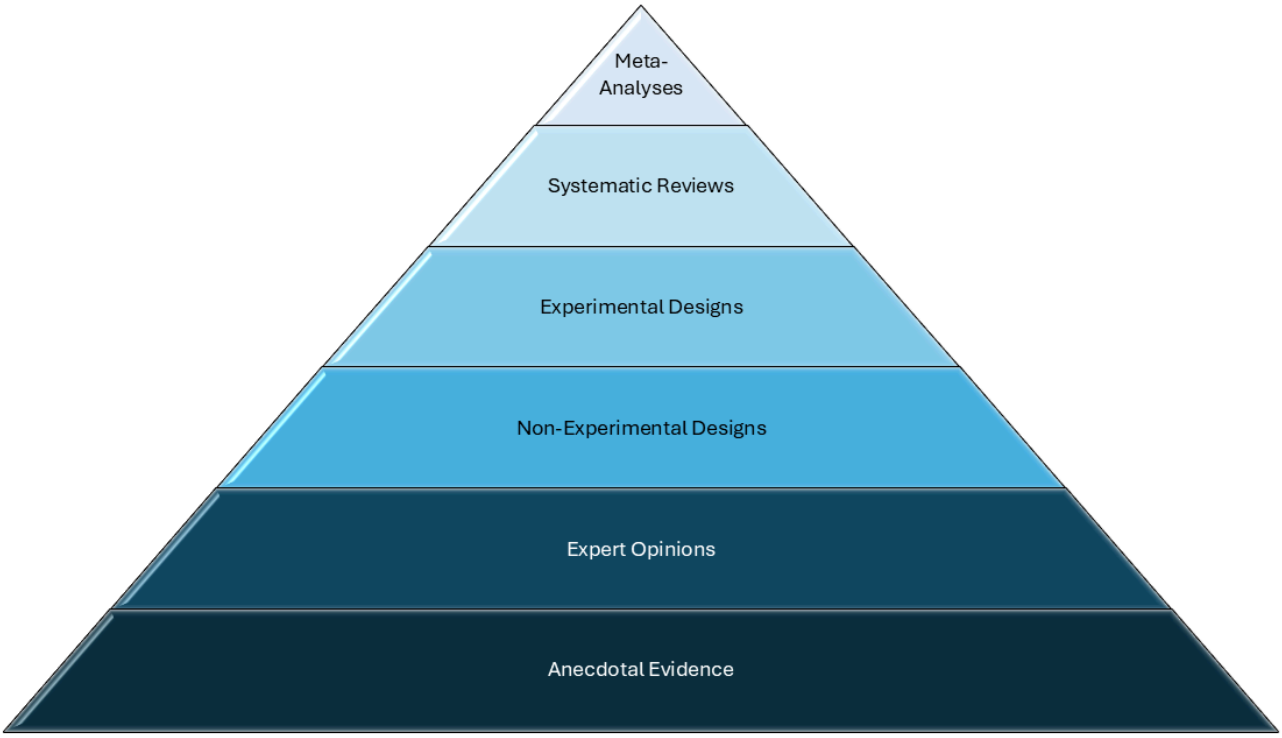


Figure 6.3.1. The hierarchy of evidence by Marc Chao is used under a CC BY-NC licence

Anecdotal Evidence

At the base of the evidence pyramid is anecdotal evidence, which consists of personal stories, individual observations, and isolated experiences. These accounts are often emotionally compelling and memorable, but they are also inherently limited in scope and prone to biases. Anecdotes can highlight interesting phenomena and raise important questions, but they cannot provide reliable answers. For example, a person might claim that taking a particular supplement cured their anxiety, but this single observation does not account for other possible explanations, such as the placebo effect, natural recovery, or coincidental changes in their environment.

Anecdotal evidence is commonly found in everyday conversations, blog posts, social media updates, online forums, and video testimonials. Platforms like Facebook, Instagram, TikTok, and Reddit are rife with personal stories shared with the intent to convince, inspire, or simply share an experience. While these anecdotes can sometimes point toward trends or raise awareness about certain issues, they lack the systematic observation, controls, and peer review necessary to establish scientific credibility. Anecdotal evidence can serve as a spark for future research, offering initial insights or highlighting gaps in existing knowledge, but it should never be used as definitive proof of cause-and-effect relationships.

Expert Opinions

Slightly higher in the hierarchy are expert opinions, which are informed perspectives offered by individuals who possess extensive experience or specialised training in a given field. Experts often have deep knowledge of a subject, and their interpretations can provide valuable insights, especially when research is limited or emerging. However, expert opinions are still subject to bias, error, and individual limitations. They rely heavily on the expert's perspective and may not always be backed by empirical evidence. For instance, a psychologist might propose a theory about memory consolidation based on years of clinical experience, but until that theory is tested empirically, it remains speculative. While expert opinions carry more weight than anecdotes, they still fall short of the rigour demanded by systematic scientific investigation.

Expert opinions are commonly found in narrative reviews, editorials, opinion pieces in academic journals, interviews with professionals in reputable publications, keynote speeches at conferences, and even podcasts or webinars featuring leading experts. In narrative reviews, experts synthesise existing knowledge on a topic and provide their interpretations, often drawing on their professional experience to highlight emerging trends or propose theoretical frameworks. While these sources are valuable for understanding current perspectives and identifying potential research directions, they should still be interpreted with caution, particularly when they lack supporting empirical data. Expert opinions serve as a useful guide, but they are most impactful when viewed as a starting point for further empirical investigation rather than definitive evidence.

Non-Experimental Designs

Moving further up the evidence ladder, we encounter non-experimental research designs, such as correlational and descriptive studies. These research designs aim to observe, measure, and classify relationships between variables without directly manipulating them. Correlational studies, for example, can reveal associations, like the observation that increased screen time is linked to poorer sleep quality, but they cannot establish causation. Descriptive studies, on the other hand, provide rich, detailed accounts of phenomena, such as case studies that explore the symptoms and behaviours of a single individual or group. While these designs are valuable for identifying patterns and generating hypotheses, they cannot definitively determine cause-and-effect relationships. They are best viewed as stepping stones toward more controlled experimental research.

Experimental Designs

At a higher level of credibility are experimental designs, which are broadly divided into two categories: non-randomised experimental designs and randomised controlled trials (RCTs). Both involve the direct manipulation of one or more independent variables while carefully controlling for confounding factors, but they differ in their level of control and the strength of causal conclusions they can provide.

Non-Randomised Experimental Designs

Also known as quasi-experimental designs, these approaches involve manipulating an independent variable and observing its effects on a dependent variable, but they do not include random assignment of participants to experimental and control groups. Instead, participants might be assigned based on pre-existing groups, convenience, or other factors outside of strict randomisation. These designs are often used in educational, clinical, or field research where randomisation is impractical, unethical, or impossible. For example, a school-based study might test a new teaching method by assigning one classroom to the intervention group and another to the control group based on existing class structures or administrative decisions. While quasi-experimental designs offer valuable insights and can establish causal relationships to some degree, the absence of random assignment introduces a higher risk of bias and confounding variables. Researchers must employ additional statistical controls and methodological rigour to account for these limitations when interpreting their findings.

Randomised Controlled Trials (RCTs)

At the pinnacle of experimental designs are randomised controlled trials (RCTs), which are widely regarded as the gold standard for establishing causal relationships. In an RCT, participants are randomly assigned to either an experimental group, which receives an intervention, or a control group, which may receive a placebo or standard treatment. This randomisation helps minimise bias and ensures that any observed differences between the groups are likely due to the intervention rather than external factors. For instance, in a study testing a new anxiety-reduction therapy, participants might be randomly assigned to either receive the therapy or undergo a placebo treatment. If the therapy group shows significantly greater improvements in anxiety symptoms, researchers can confidently attribute the effect to the intervention itself. RCTs are especially valuable in medical and clinical psychology research, where precision and reliability are critical for informing evidence-based practices.

Systematic Reviews

Above experimental studies sit systematic reviews, which represent a synthesis of evidence from multiple studies addressing the same research question. Researchers conducting a systematic review follow a rigorous and transparent process to search for, evaluate, and summarise all available evidence on a particular topic. This approach minimises bias by including studies with varying results and methodologies, offering a more comprehensive and balanced view of the existing evidence. For example, a systematic review of research on the effectiveness of cognitive-behavioural therapy (CBT) for treating depression might include dozens of studies from different countries, populations, and clinical settings, ultimately painting a clearer picture of CBT's overall effectiveness.

Meta-Analyses

One step higher in reliability is the meta-analysis, which goes beyond merely summarising studies by using statistical techniques to combine data from multiple studies into a single quantitative estimate of an effect. Meta-analyses not only pool results from individual studies but also weigh them based on factors like sample size, study quality, and statistical significance. This approach allows researchers to identify patterns, measure effect sizes, and account for inconsistencies across studies. For example, a meta-analysis of studies on mindfulness-based stress reduction (MBSR) might reveal not only that MBSR is effective but also how its effects vary depending on factors such as participant age or duration of the intervention. Because of their statistical rigour and ability to aggregate large amounts of data, meta-analyses are considered one of the most robust forms of evidence available.

At the pinnacle of the evidence hierarchy are Cochrane meta-analyses, named after the Cochrane Collaboration, a global network of researchers committed to producing high-quality, evidence-based reviews. Cochrane meta-analyses are held to exceptionally high standards of methodological transparency, reproducibility, and objectivity. Each review undergoes a meticulous process of study selection, quality assessment, and statistical analysis. These reviews are frequently updated to include the latest research, ensuring that their conclusions remain current and accurate. For example, a Cochrane meta-analysis on the efficacy of antidepressants for treating major depressive disorder would provide one of the most authoritative summaries of the available evidence, making it a trusted resource for clinicians, policymakers, and researchers alike.

Incorporating Lived Experience into Scientific Research Syntheses

This section of the book has explained the various sources of evidence for understanding the social world. It also illustrated the credibility and reliability of each source of evidence when building arguments, making decisions, or advancing knowledge. While systematic reviews and meta-analyses are placed at the top of the hierarchy of evidence in Figure 6.3.1, it must be noted that in these methods of research syntheses, the identification of research questions, review of the extant literature (existing published research), generating interpretations, and identifying implications and recommendations for research and policy making are typically done without the involvement of the individuals and communities with lived experience (Beames et al., 2021; Grindell et al., 2022). Incorporating the views and perspectives of target groups with lived experience (for example, people living with various health conditions, or those from marginalised communities) as part of the scientific research process, including research syntheses, can ensure that the research design, data collection and analysis are centred on their needs and priorities. This, in turn, can ensure that findings and recommendations generated from scientific research can benefit these target groups the most.

These co-creation, co-design, and co-production principles are integral to participatory research designs, where academics and researchers actively collaborate with patients, consumers, and other relevant stakeholders throughout the knowledge generation process (Grindell et al., 2022; Vargas et al., 2022). For example, academics and researchers can involve people with mental health issues as part of the research design process, where they collaborate on defining the appropriate research questions that drive the research and co-designing the study's data collection tools (e.g., survey questions and interview guides). This ensures that the research being conducted is grounded in real-world needs through the inclusion of voices and perspectives from those with lived experience. Similarly, in research syntheses, individuals with mental health issues can be involved in the systematic review process by providing input on the review questions, as well as reflecting and commenting on the findings generated from the review (Beames et al., 2021). Such integrative review methods can provide additional insights into other unexplored areas within mental health and also lead to the timely development and implementation of contextually-appropriate interventions that can benefit other people with mental health issues.

References

Beames, J. R., Kikas, K., O'Gradey-Lee, M., Gale, N., Werner-Seidler, A., Boydell, K. M., & Hudson, J. L. (2021). A new normal: Integrating lived experience into scientific data syntheses. *Frontiers in Psychiatry*, *12*, 1-4. https://doi.org/10.3389/fpsyt.2021.763005

Grindell, C., Coates, E., Croot, L., & O'Cathain, A. (2022). The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: A systematic review. *BMC Health Services Research*, *22*(1), 1-26. https://doi.org/10.1186/s12913-022-08079-y

Vargas, C., Whelan, J., Brimblecombe, J., & Allendera, S. (2022). Co-creation, co-design and co-production for public health: A perspective on definitions and distinctions. *Public Health Research & Practice*, *32*(2), 1-7. https://doi.org/10.17061/phrp3222211

6.4. GENERATING RESEARCH QUESTIONS AND HYPOTHESES

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Formulating an empirically testable research question is a crucial step in the scientific process. It transforms a general research idea into a focused inquiry that can be investigated systematically. At its core, an empirically testable question must involve observable and measurable variables, either focusing on a single variable or exploring the relationship between multiple variables. While this process may seem daunting at first, as though experienced researchers pluck compelling questions out of thin air, it is, in reality, the result of strategic thinking, persistence, and familiarity with the research literature.

One effective approach to generating research questions is to examine the discussion sections of recent academic articles. The discussion section is where researchers interpret their results, relate them to past studies, and suggest directions for future research. These suggestions often highlight unanswered questions, methodological limitations, or intriguing findings that warrant further exploration. Because these future directions have already been flagged by experienced researchers as meaningful and important, they offer fertile ground for developing your own research questions.

Beyond reviewing existing studies, you can also generate research questions by starting with a specific behaviour or psychological characteristic and framing it as a variable. For example, you might ask: How many words do people speak in a day? How accurate are people's memories of traumatic events? What percentage of adults experience chronic anxiety? If these questions have not yet been thoroughly explored, as you will discover during your literature review, they might represent valuable research opportunities.

If a variable has already been studied extensively, the next step is to consider relationships between variables. For instance, you might ask what factors cause a particular behaviour or psychological characteristic, what consequences it might have, or how it varies across different people or situations. If you are interested in talkativeness, you might ask whether family size influences how much people talk, or whether same-sex social groups foster more conversation than mixed-sex groups. Each potential relationship represents a unique research question that could contribute to the broader scientific understanding of your topic.

However, encountering a question that has already been answered by previous research does not mean you should abandon it. Instead, consider refining the question to offer a fresh perspective or address a gap in the literature. Are there alternative ways to define or measure the variables in question? Are there specific populations where the relationship might be stronger or weaker? Could situational factors influence the outcome in meaningful ways? For example, while previous research suggests men and women speak about the same number of words per day, you might refine the question by focusing on whether this finding holds true across different age groups or cultural contexts. Through this process, even well-explored topics can yield new avenues for investigation.

Evaluating Research Questions

Generating a list of potential research questions is only the beginning. Researchers must carefully evaluate each question to determine which are worth pursuing. Two essential criteria guide this evaluation: interestingness and feasibility.

Interestingness

A research question's interestingness is not about whether it fascinates you personally, but whether it holds broader relevance for the scientific community. Several factors determine this.

First, a research question is interesting if its answer is genuinely in doubt. Questions that have already been conclusively answered through prior research are no longer compelling subjects for new investigation. However, if reasonable arguments can be made for multiple potential answers, the question becomes much more engaging. For example, the question of whether women are more talkative than men is intriguing because plausible arguments exist on both sides: stereotypes suggest women are more talkative, but evidence shows little difference in verbal abilities between genders.

Second, a question is interesting if it fills a gap in the existing literature. Even if a question has not been answered empirically, it must feel like a natural and meaningful extension of what is already known. For example, asking whether taking notes by hand improves academic performance naturally follows from research showing the cognitive benefits of deeper information processing.

Finally, a research question gains significance if it has practical implications. Questions that address real-world problems or inform practical decision-making are often considered more valuable. For instance, exploring whether cell phone use impairs driving performance carries meaningful consequences for public safety and policy-making.

Feasibility

A research question might be theoretically fascinating, but if it cannot realistically be answered given your resources, expertise, or timeline, it is not worth pursuing. Feasibility depends on several factors, including time, funding, access to equipment, technical skills, and availability of participants.

For instance, a large-scale longitudinal study tracking participants over decades would require significant time and funding, resources typically unavailable to a single student researcher. Similarly, a neuroimaging study involving advanced brain-scanning technologies may not be feasible without access to specialised labs and training.

However, feasibility does not mean compromising on the quality of your study. Many impactful studies are relatively simple and resource-friendly, relying on university student samples or straightforward observational tasks. Even small-scale studies can yield meaningful contributions if they are well-designed and methodologically sound.

When designing your study, it is often wise to borrow methods from existing research. If previous studies have successfully manipulated participants' moods by offering compliments, for example, adopting this approach is both practical and methodologically consistent. Not only does this increase the likelihood of success, but it also ensures your findings are easier to compare with existing literature.

Theories and Hypotheses

Understanding the difference between a theory and a hypothesis is essential for conducting meaningful scientific research. While these two terms are often used interchangeably in everyday conversation, they have distinct meanings in the realm of science. A theory is a coherent and systematic explanation of one or more phenomena, built upon established evidence and reasoning. It serves as a framework for understanding and predicting outcomes. Theories often introduce abstract concepts, relationships, and processes that go beyond the observable data.

For example, Zajonc's theory of social facilitation and social inhibition (1965) suggests that being observed by others during a task creates a state of physiological arousal. This arousal, in turn, enhances the performance of well-practised tasks (social facilitation) but impairs performance on unfamiliar or complex tasks (social inhibition). The theory introduces terms like arousal and dominant response, which are not directly observable but serve as essential constructs for explaining observed behaviour patterns. Such theoretical constructs provide a foundation for generating specific hypotheses and guiding further research.

It is important to note that in science, the term theory does not imply uncertainty or guesswork, as it often does in everyday language. A scientific theory can be extensively tested, well-supported, and widely accepted by the scientific community. For instance, the theory of evolution by natural selection and the germ theory of disease are both referred to as theories, not because they are speculative, but because they provide comprehensive explanations for large sets of observed phenomena. These theories are supported by vast amounts of empirical evidence and continue to guide scientific discovery.

In contrast, a hypothesis is a specific, testable prediction about what should be observed if a theory is accurate. Hypotheses are narrower in scope and focus on particular aspects of a theory or phenomena. They are formulated based on existing evidence, logical reasoning, or theoretical frameworks and are often stated in ways that allow them to be tested empirically. For example, based on Zajonc's drive theory, one might hypothesise: If drive theory is correct, then cockroaches should run faster through a straight runway but slower through a branching runway when other cockroaches are present.

However, not all hypotheses are derived from existing theories. In some cases, researchers generate atheoretical hypotheses, which arise from observations or preliminary data without being directly tied to an overarching theory. For example, if researchers notice an unexpected behavioural pattern during preliminary observations, they might develop a hypothesis to investigate that pattern further. Over time, a broader theory might emerge from a collection of related hypotheses and findings.

Hypotheses often take the form of if-then statements, establishing a clear relationship between variables. For example, if expressive writing helps people habituate to negative emotions, then writing about traumatic experiences should reduce emotional distress more effectively than writing about positive experiences. Even when stated as declarative sentences, hypotheses can always be rephrased as research questions, such as "Does expressive writing about traumatic experiences reduce emotional distress more than writing about positive experiences?"

Deriving Hypotheses from Theories

The process of generating hypotheses from theories typically begins with identifying a research question. Researchers can then ask whether any existing theory provides a potential answer to that question. For instance, if a researcher wonders whether writing about positive life events has the same psychological benefits as writing about traumatic events, they might turn to habituation theory. According to habituation theory, emotional benefits arise from repeated exposure to negative thoughts and feelings, which reduces their emotional impact over time. If this theory is correct, writing about positive experiences should not yield the same benefits as writing about traumatic experiences because positive events do not evoke distress that requires habituation.

Another way to derive hypotheses from theories is to focus on specific components or mechanisms within the theory that have not yet been directly observed or tested. For example, a researcher could examine whether emotional habituation happens gradually across multiple expressive writing sessions by measuring participants' distress levels after each session.

Among the most valuable hypotheses are those that can distinguish between competing theories. For example, Norbert Schwarz and his colleagues (1991) investigated two competing theories about how people judge their assertiveness. One theory proposed that people base their self-judgements on the number of relevant examples they can recall, while the other theory suggested that judgements are based on how easily those examples come to mind. To test these theories, participants were asked to recall either six (easy) or twelve (difficult) examples of their assertive behaviour and then rate their overall assertiveness. The first theory predicted that recalling more examples would lead to higher assertiveness ratings, while the second theory predicted that ease of recall would play a more significant role. The results supported the ease-of-retrieval theory, demonstrating the value of crafting hypotheses that pit one theory against another.

Theory Testing and the Hypothetico-Deductive Method

The process of testing theories follows a method known as the hypothetico-deductive approach. Researchers start with an existing theory or construct one based on observed phenomena. From this theory, they derive specific hypotheses, which are predictions about what should occur under certain conditions if the theory is accurate. They then design and conduct empirical studies to test these hypotheses. Based on the results, the theory is either supported, refined, or revised. This cyclical process, as shown in Figure 6.4.1, is essential for advancing scientific understanding, as each iteration builds upon the findings of previous research.

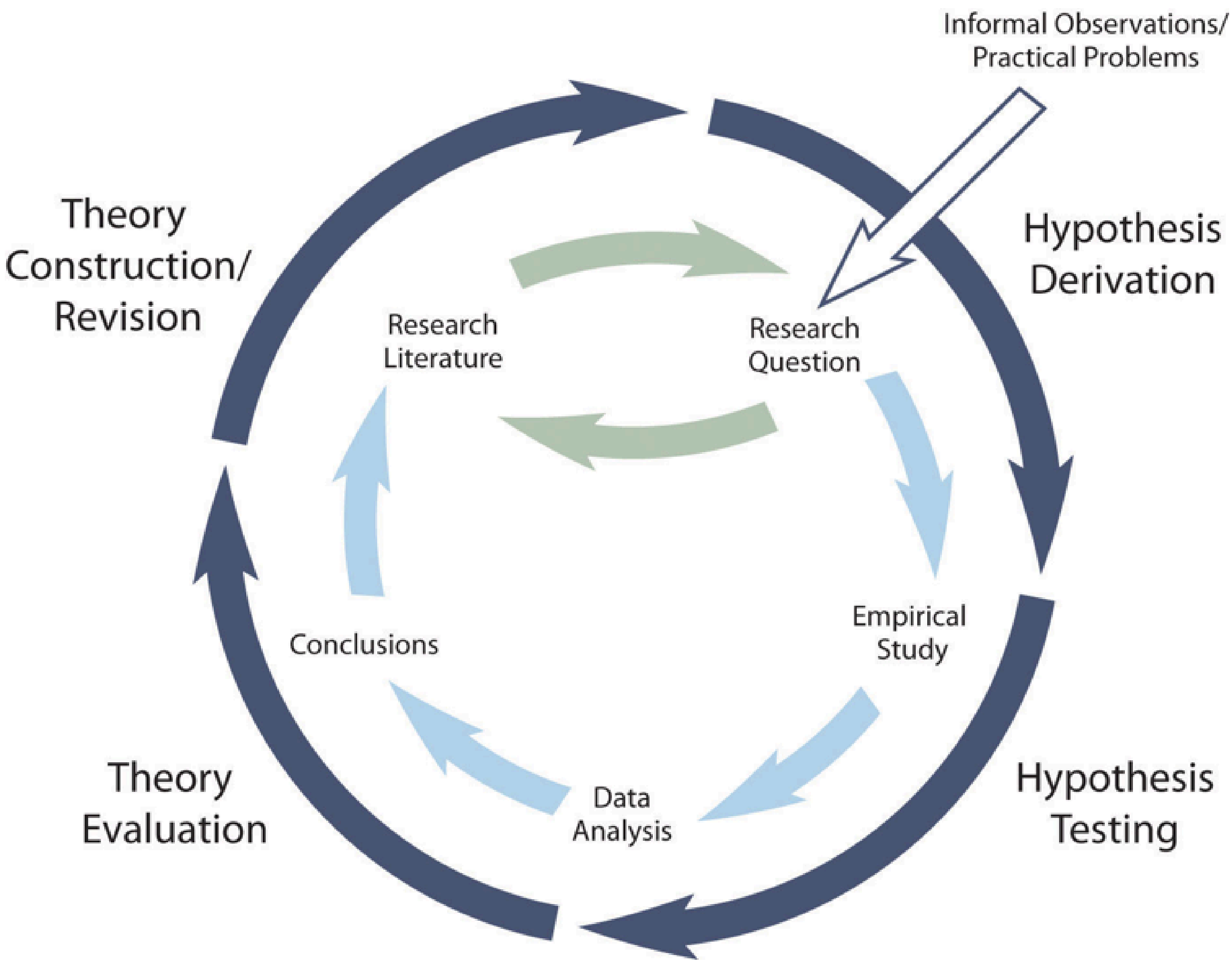


Figure 6.4.1. Hypothetico-deductive method combined with the general model of scientific research in psychology. Together, they form a model of theoretically motivated research by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

A classic example of this approach comes from Zajonc's research on social facilitation and inhibition. After developing drive theory, Zajonc hypothesised that cockroaches would perform better on simple tasks and worse on complex tasks when observed by others. His experiments confirmed these predictions, providing strong empirical support for his theory. This iterative process demonstrates how theory-driven research contributes to scientific progress by refining and expanding theoretical frameworks.

Incorporating Theory into Your Research

Incorporating theory into your research enhances its significance and clarity. There are two primary ways researchers typically use theories in their work. The first approach involves conducting a study to answer a research question and then using one or more theories to interpret the results. This approach is particularly useful for applied research or when existing theories do not directly address the question at hand. The second approach involves deriving a hypothesis from an existing theory, testing that hypothesis through an empirical study, and then evaluating or refining the theory based on the results.

Using established theories not only strengthens the foundation of your research but also situates your work within the broader scientific dialogue. Psychological theories are the result of decades of research and represent collective knowledge about human behaviour and mental processes. By aligning your research with these theoretical frameworks, you ensure that your findings contribute meaningfully to the scientific community.

Characteristics of a Good Hypothesis

A strong hypothesis possesses three key characteristics: testability, logical reasoning, and positivity.

First, a hypothesis must be testable and falsifiable. This means it must be possible to gather empirical evidence that could disprove the hypothesis if it is incorrect. If a hypothesis cannot be tested or proven false, it falls outside the realm of scientific inquiry.

Second, a hypothesis must be logical. It should be informed by existing theories, observations, or empirical data and should follow a clear line of reasoning. Hypotheses are not random guesses; they emerge from a structured thought process that connects prior knowledge to new questions.

Finally, a hypothesis should be positive. It should make a statement about the existence of a relationship or effect rather than the absence of one. Scientists begin with the assumption that no effect exists (the null hypothesis) and then look for evidence to reject this assumption in favour of an alternative hypothesis.

By crafting hypotheses that are testable, logical, and positive, researchers create clear, focused questions that can be systematically investigated, ultimately advancing our understanding of complex phenomena.

References

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

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

6.5. DESIGNING A RESEARCH STUDY

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Identifying and Defining the Variables and Population

Variables and Operational Definitions

At the heart of every psychological study are variables, which represent the core elements researchers seek to measure, manipulate, or observe. A variable refers to any characteristic, behaviour, or condition that can vary across individuals or situations. For example, a person's height, age, or level of anxiety are all variables because they differ from one individual to another. Similarly, a person's chosen academic major or cultural background also qualifies as variables, as they are qualities that can differ across people.

Variables are typically categorised into two main types: quantitative variables and categorical variables. Quantitative variables represent measurable quantities, often expressed numerically. For example, a person's height (in centimetres), the number of hours they sleep per night, or their score on an anxiety questionnaire are all quantitative variables. On the other hand, categorical variables represent qualities or classifications, often assigned as labels rather than numbers. Examples of categorical variables include a person's nationality, their gender identity, or their preferred type of music.

Once researchers identify the variables they want to study, they must define them in a way that allows for systematic measurement. This is done through operational definitions. An operational definition specifies precisely how a variable will be measured or manipulated in the context of a study. This step is critical because many psychological variables, such as anxiety, depression, or happiness, are abstract concepts that cannot be directly observed or measured. For example, depression might be operationally defined in several ways: as a participant's score on the Beck Depression Inventory, the number of depressive symptoms they report, or whether they have been clinically diagnosed with major depressive disorder.

Operational definitions not only allow researchers to measure abstract concepts but also ensure consistency across studies. If a particular operational definition has been widely used in previous research, it is generally a good idea to adopt it. This helps align the new study with the existing body of research, allowing for better comparisons and interpretations.

Sampling and Measurement

Once the variables and their operational definitions are established, researchers must consider the population they aim to study. In psychology, the population refers to the entire group of individuals that the researcher wants to draw conclusions about. Depending on the study's goals, this population might be all teenagers in the United States, individuals diagnosed with anxiety disorders, or even all human beings.

However, studying an entire population is almost always impractical. Instead, researchers select a sample, which is a smaller subset of the population. The goal is to ensure that this sample is representative of the larger population, meaning it reflects the key characteristics of the group being studied.

One approach to obtaining a sample is simple random sampling, where every member of the population has an equal chance of being selected. For example, if a researcher wants to study voting behaviour, they might randomly select names from a list of registered voters. While random sampling offers strong advantages in terms of representativeness, it is often impractical in psychological research. Populations like "all teenagers in the United States" or "children with autism" are challenging to define and access comprehensively.

Instead, most psychological research relies on convenience sampling, where participants are selected based on their availability and willingness to participate. University psychology students are often included in such samples because they are readily accessible to academic researchers. However, convenience sampling carries the risk of introducing sampling bias, where the sample might not fully represent the broader population. Researchers must carefully consider this limitation when interpreting and generalising their results.

Experimental vs. Non-Experimental Research

After identifying the variables and defining the population, researchers must decide how they will approach data collection. One of the most fundamental distinctions in psychological research lies between experimental and non-experimental approaches.

Experimental Research

Experimental research is designed to test causal relationships between variables. This approach involves the manipulation of an independent variable and the measurement of its effect on a dependent variable while controlling for extraneous variables. For example, if a researcher wants to test whether sleep deprivation affects memory performance, they might manipulate sleep (e.g., full night of sleep vs. no sleep) and measure participants' memory recall performance.

In experimental research, the independent variable (IV) is the factor manipulated by the researcher, while the dependent variable (DV) is the outcome being measured. Researchers must also carefully address confounding variables, which are factors that unintentionally vary alongside the independent variable and might offer alternative explanations for observed results. For instance, if participants in the sleep deprivation group are also exposed to loud noises, it would be unclear whether poor memory performance was caused by lack of sleep or by the noise.

Non-Experimental Research

In non-experimental research, researchers observe and measure variables as they naturally occur, without manipulating them. This approach is useful for describing phenomena, identifying relationships between variables, and making predictions. For example, a researcher might examine the correlation between social media usage and self-esteem by measuring how much time participants spend on social media and their scores on a self-esteem scale. While non-experimental research can identify patterns and associations, it cannot establish causal relationships because there is no manipulation or control over variables.

It is important to note that non-experimental research is still scientific. It can effectively fulfil two key goals of science: description and prediction. However, it cannot address the third goal: explanation, because causality cannot be established without experimental control.

Laboratory vs. Field Research

Another distinction in research design lies between laboratory research and field research. Each approach offers unique advantages and limitations, often trading off between internal validity and external validity.

Laboratory Research

Laboratory studies are conducted in controlled, artificial environments where researchers can manipulate variables with precision and minimise extraneous factors. The strength of laboratory research lies in its high internal validity, which refers to the confidence that observed effects are genuinely caused by the manipulated independent variable. For example, in a lab study examining the effects of caffeine on concentration, researchers can carefully control participants' caffeine intake, monitor their performance on cognitive tasks, and minimise external distractions.

However, laboratory studies often lack external validity, meaning their findings may not generalise well to real-world situations due to the artificial nature of the environment.

Field Research

Field studies, in contrast, are conducted in natural settings, such as workplaces, schools, or public spaces. This approach excels in external validity, as behaviours and responses are observed in real-world conditions. For example, a study on teamwork dynamics conducted in an actual corporate office might provide insights that laboratory simulations cannot capture.

However, field research often suffers from lower internal validity because researchers have less control over extraneous variables. Unexpected factors like interruptions, environmental noise, or individual differences can complicate data interpretation.

Interestingly, field experiments combine elements of both approaches. Researchers manipulate an independent variable in a natural setting while still attempting to control for extraneous variables. When done carefully, field experiments can achieve both high internal and external validity, offering robust and generalisable findings.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

6.6. ANALYSING THE DATA

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Once a study has been conducted and the data collected, researchers must systematically analyse the data to draw meaningful conclusions. This stage is critical because raw data, no matter how extensive, do not speak for themselves. Data analysis involves applying statistical techniques to identify patterns, relationships, and trends within the data. Typically, researchers use two primary types of statistics: descriptive statistics and inferential statistics. Together, these approaches help summarise the data, test hypotheses, and determine whether the results can be generalised to a larger population.

Descriptive Statistics

Descriptive statistics are used to summarise, organise, and simplify data so that they can be more easily interpreted. They provide a clear picture of what the data look like and allow researchers to highlight key patterns. Descriptive statistics typically fall into three major categories: measures of central tendency, measures of dispersion, and correlation coefficients.

Measures of Central Tendency

These statistics describe the centre or average value of a data set and give researchers an idea of the "typical" response within the sample. The three main measures of central tendency are:

- **Mean:** The arithmetic average of a set of scores. It is calculated by adding all the scores together and dividing by the number of scores.
- **Median:** The middle score in a dataset when the scores are arranged in ascending or descending order.
- **Mode:** The most frequently occurring score in a dataset.

For example, if researchers measure how many hours a group of students sleep per night, the mean would give the average number of hours, the median would show the midpoint value, and the mode would identify the most common number of hours reported.

Measures of Dispersion

While measures of central tendency show the average or typical value, measures of dispersion indicate how spread out the data are around that central point. These include:

- **Range:** The difference between the highest and lowest scores.
- **Standard Deviation:** A more sophisticated measure that indicates how far, on average, each score deviates from the mean.
- **Variance:** The square of the standard deviation, providing another measure of spread, though less commonly interpreted directly.

For instance, if two classrooms report an average test score of 85, but one classroom has a standard deviation of 2 while the other has a standard deviation of 15, the latter classroom shows far greater variability in student performance.

Correlation Coefficients

In non-experimental research, researchers often seek to identify relationships between two variables rather than comparing groups. The correlation coefficient measures both the strength and direction of these relationships, ranging from -1.00 to +1.00:

- A positive correlation means that as one variable increases, the other also increases (e.g., height and weight).
- A negative correlation means that as one variable increases, the other decreases (e.g., stress and happiness).
- A correlation coefficient close to 0 indicates no relationship between the variables.

For example, if researchers observe a correlation of +0.75 between sleep duration and cognitive performance, it suggests a strong positive relationship, where better sleep is associated with better cognitive outcomes.

Descriptive statistics serve as the foundation for understanding the dataset, preparing it for more complex statistical analysis, and communicating the findings in a clear and accessible way.

Inferential Statistics

While descriptive statistics summarise what happened within the sample, inferential statistics allow researchers to draw conclusions about the broader population based on sample data. This process is crucial because most psychological studies rely on samples rather than entire populations.

Inferential statistics enable researchers to test hypotheses and determine whether the observed effects in their data are statistically significant. In other words, whether they are unlikely to have occurred by chance.

Statistical Significance and Probability

Statistical significance is determined using a *p*-value. This is a probability value that indicates the likelihood of obtaining the observed results if the null hypothesis (the assumption that there is no real effect or relationship) were true. In most research, a *p*-value of less than 0.05 (5%) is considered statistically significant. This means there is less than a 5% chance that the observed effect occurred randomly.

For example, if a study finds that a new anxiety treatment significantly reduces symptoms compared to a placebo, and the *p*-value is less than 0.05, researchers can conclude that the effect is unlikely to be due to chance.

The Role of Probability and Error

It is important to note that inferential statistics are probabilistic and never provide absolute certainty. Instead, they offer confidence levels about whether an observed effect reflects a real relationship in the population. However, this probabilistic nature opens the door to potential errors:

- **Type I Error (False Positive):** This occurs when researchers conclude that an effect exists when it actually does not. For example, they might conclude that a drug improves memory when the observed results were purely due to chance. The 5% significance threshold helps minimise this risk but does not eliminate it entirely.
- **Type II Error (False Negative):** This happens when researchers fail to detect an effect that actually exists. For instance, they might conclude that a treatment has no impact when it genuinely does, perhaps because the sample size was too small or the statistical power was inadequate.

Researchers aim to strike a balance between minimising Type I and Type II errors, often adjusting sample sizes, significance thresholds, and statistical techniques to ensure their conclusions are as reliable as possible.

Drawing Conclusions from Statistical Analyses

Once researchers have completed their statistical analyses, they must carefully interpret their results. Did the findings support the hypothesis? Were there unexpected patterns? Do the results align with or contradict previous research?

Statistical Significance vs. Practical Significance

While statistical significance indicates whether an effect is unlikely due to chance, practical significance considers whether the effect is meaningful in real-world terms. For example, if a study finds that a drug reduces anxiety scores by 0.5 points on a 100-point scale, the result might be statistically significant but not practically meaningful.

Replicability and Transparency

To strengthen confidence in their findings, researchers often conduct replication studies by repeating the experiment under similar conditions to see if the same results emerge. They also share their data, methods, and analyses transparently, enabling other scientists to verify or challenge their conclusions.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

6.7. DRAWING CONCLUSIONS AND REPORTING THE RESULTS

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Drawing Conclusions

Scientific research is inherently probabilistic, meaning its findings are subject to uncertainty and the possibility of error. Because of this, a single study can rarely, if ever, offer absolute certainty about a theory. Instead of aiming to "prove" theories, scientists focus on supporting, refuting, or refining them based on patterns of evidence that emerge from empirical studies.

When the results of a study are statistically significant and align with the predictions made by a hypothesis, researchers can conclude that the findings support the underlying theory. In such cases, the theory not only made an accurate prediction, but it now accounts for a new phenomenon supported by empirical data. Conversely, when the results fail to support the hypothesis, the theory is weakened. The inaccurate prediction suggests a gap in the theory's explanatory power, highlighting a phenomenon it does not fully address.

However, this process is not as straightforward as it may initially seem. A confirmed hypothesis can strengthen a theory, but it cannot definitively prove it. Scientists are careful to avoid using the word "prove" when discussing their theories for several reasons. First, statistically significant results could still stem from a Type I error, which is a false positive where the observed effect occurred by chance. Second, multiple theories may predict the same hypothesis, meaning that confirming a hypothesis could equally support all those competing theories. Finally, the problem of induction, a well-known philosophical issue, underscores that no number of confirming observations can eliminate the possibility of encountering a disconfirming one in the future. For instance, observing countless white swans cannot rule out the existence of a single black swan. Because of these limitations, even widely accepted theories remain subject to revision as new evidence emerges.

Disconfirmed hypotheses also introduce their own complexities. According to the strict hypothetico-deductive method, if a hypothesis derived from a theory is not supported, it logically suggests that the theory itself is flawed. In formal logic, if the premise "If A, then B" is paired with the observation "not B", the conclusion must be "not A". In practice, however, scientists rarely discard a theory after a single disconfirmed hypothesis. There are several reasons for this caution. A failed hypothesis could result from a Type II error, where a real effect was missed due to insufficient statistical power or a small sample size. Alternatively, the research design might have been flawed; for example, the independent variable may not have been manipulated effectively, or the dependent variable may not have been measured accurately.

Sometimes, disconfirmation reveals a previously overlooked assumption within the theory. For example, if Zajonc had failed to find evidence for social facilitation in cockroaches, he might have concluded that the drive theory still holds but applies only to organisms with more complex nervous systems. In such cases, researchers refine or adjust their theories rather than discarding them entirely. However, repeated disconfirmations across multiple studies, especially with improved methodologies, eventually necessitate abandoning the theory in favour of one better supported by evidence.

The key takeaway is that science deals in evidence, not proof. Because all studies carry some level of error and uncertainty, scientific conclusions are always open to refinement, reinterpretation, or rejection in light of new data.

Reporting the Results

The final step in the scientific research process is communicating the findings to the broader scientific community and, in some cases, to the public. Transparent and thorough reporting is essential for advancing knowledge, fostering collaboration, and enabling others to replicate or build upon the research.

One of the most prestigious and rigorous methods for sharing research findings is through peer-reviewed journal articles. These articles are submitted to academic journals and undergo a thorough peer-review process, where other experts in the field critically evaluate the study's methodology, analysis, and conclusions. If the research meets the journal's standards, it is accepted for publication. In psychology, these articles are usually written in accordance with the American Psychological Association (APA) style, a standardised format that ensures clarity, consistency, and proper attribution of sources.

Another common platform for sharing research findings is through book chapters in edited volumes. These chapters are typically contributions from various researchers, each focusing on a specific aspect of a broader topic. While some edited volumes undergo peer review, others may rely on the expertise of the editors to ensure quality. Book chapters allow researchers to delve deeper into their findings and explore theoretical implications in a more extended format than a journal article might allow.

In addition to written publications, many researchers choose to present their findings at academic conferences. Conferences provide an opportunity for direct engagement with peers, fostering discussion, feedback, and potential collaborations. Presentations at conferences generally take one of two forms: oral presentations or poster presentations.

Oral presentations involve standing in front of an audience and delivering a talk that typically lasts between

10 minutes and an hour, followed by a question-and-answer session. These presentations allow researchers to highlight their most significant findings and clarify complex points in real time.

Poster presentations, on the other hand, consist of summarising the study on a large printed or digital poster. Posters typically include key sections such as purpose, methodology, results, and conclusions. Researchers stand by their posters during designated sessions, answering questions and engaging in discussions with attendees who stop to learn more about their work. Poster presentations are especially valuable for receiving constructive feedback before submitting a manuscript for peer-reviewed publication.

Beyond academic channels, researchers may also disseminate their findings through public talks, blog posts, and media interviews, depending on the nature of the research and its relevance to broader audiences. Sharing research publicly helps bridge the gap between academia and society, ensuring that scientific insights contribute to public knowledge, policy changes, and real-world applications.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

6.8. SUMMARY

By Marc Chao

Summary

Scientific research in psychology operates as a dynamic, iterative cycle that begins with formulating research questions and progresses through empirical studies, data analysis, conclusion drawing, and sharing findings. This process is fuelled by the research literature, which serves as a repository of established knowledge and a source of inspiration for new questions. Studies like Mehl's investigation into talkativeness and research on cell phone use while driving illustrate how systematic inquiry addresses stereotypes and practical problems, contributing to both academic understanding and societal change. By refining knowledge through observation, experimentation, and application, psychology advances its grasp of human behaviour and addresses real-world challenges.

The development of strong research questions blends creativity and structure, often drawing inspiration from informal observations, practical issues, or gaps in existing literature. Reviewing the research literature is crucial for refining questions and situating them within the broader scientific context. Peer-reviewed journals, databases like PsycINFO, and strategies such as citation tracking ensure researchers access high-quality, relevant sources. This process ensures that new investigations are grounded in existing knowledge and contribute meaningfully to the field.

Scientific evidence follows a hierarchy based on its reliability and utility. While anecdotal evidence and expert opinions offer initial insights, non-experimental designs and experimental research provide increasingly robust findings. Randomised controlled trials (RCTs) offer the strongest causal evidence among experimental designs. At the top of the hierarchy, systematic reviews and meta-analyses synthesise data to produce comprehensive conclusions, with Cochrane meta-analyses exemplifying the highest standard of rigour. This hierarchy underscores the importance of using reliable evidence to inform research and practice.

Formulating empirically testable research questions requires focusing on observable and

measurable variables, often informed by academic literature, practical problems, or informal observations. Researchers refine questions to explore variable relationships or address gaps, evaluating them for interestingness and feasibility. Hypotheses derived from theories provide specific, testable predictions, and the hypothetico-deductive method allows findings to refine or challenge theoretical frameworks. Effective hypotheses are testable, logically grounded, and positively framed, facilitating systematic investigation and meaningful scientific contributions.

Psychological research systematically identifies and defines variables and populations to study human behaviour. Variables, whether quantitative (e.g., sleep hours) or categorical (e.g., nationality), are operationally defined for measurable consistency. Due to practical constraints, researchers rely on representative samples, often using convenience sampling. Experimental research manipulates variables to test causality, while non-experimental designs explore relationships. Laboratory research prioritises internal validity, and field research emphasises external validity. Field experiments blend these strengths, ensuring credibility and real-world relevance.

Data analysis transforms raw data into meaningful conclusions through statistical techniques. Descriptive statistics summarise data with measures of central tendency, dispersion, and correlation, offering an overview of patterns. Inferential statistics generalise findings to broader populations, testing hypotheses and determining statistical significance through tools like *p*-values. Researchers also evaluate practical significance to assess real-world impact while striving to minimise errors and ensure replicability and transparency. This systematic approach ensures robust conclusions that advance scientific understanding.

Scientific research focuses on refining, supporting, or refuting theories, acknowledging the inherent uncertainty of all studies. Confirmed hypotheses strengthen but do not prove theories, while disconfirmed hypotheses prompt refinement or highlight methodological gaps. Over time, repeated disconfirmations may lead to replacing a theory with a better-supported alternative. Findings are disseminated through peer-reviewed journals, academic conferences, and public platforms, ensuring transparency, collaboration, and societal relevance. This iterative process drives the advancement and application of scientific knowledge in meaningful ways.

CHAPTER 7: RESEARCH ETHICS

Research ethics form the foundation of responsible scientific inquiry, guiding researchers in balancing the pursuit of knowledge with the protection of participants' rights and well-being. Ethical considerations in psychology extend beyond simple rules; they represent a framework for thoughtful decision-making throughout every stage of the research process. These principles not only safeguard participants from harm but also ensure the credibility, reliability, and integrity of scientific findings.

This chapter examines the ethical landscape of psychological research, exploring key moral principles such as weighing risks against benefits, acting with responsibility and integrity, seeking justice, and respecting participants' rights and dignity. It traces the historical development of research ethics, including the influence of landmark documents like the Nuremberg Code, the Declaration of Helsinki, and the Belmont Report, which collectively laid the groundwork for contemporary ethical standards.

The chapter also focuses on the American Psychological Association (APA) Ethical Principles of Psychologists and Code of Conduct, particularly Standard 8, which addresses research and publication. This code offers clear guidelines on essential practices, including obtaining informed consent, minimising risks, handling deception responsibly, ensuring thorough debriefing, and protecting participant confidentiality. The humane treatment of nonhuman animal subjects and the importance of scholarly integrity, including avoiding plagiarism and fabricating data, are also emphasised.

Furthermore, the chapter highlights the practical responsibilities of researchers, from securing institutional approval through an Institutional Review Board (IRB) to navigating challenges during data collection and reporting findings transparently. It underscores the importance of identifying and minimising risks, justifying the use of deception, and carefully weighing the potential benefits of research against its risks.

Ultimately, the chapter reinforces that ethical research is not merely about compliance with guidelines; it requires a deep commitment to fairness, transparency, and respect for all individuals involved. Researchers must remain vigilant and thoughtful, continuously reflecting on the ethical implications of their work and ensuring that their pursuit of scientific knowledge aligns with the highest moral standards.

Learning Objectives

By the end of this chapter, you should be able to:

- **Define ethics and its role in research:** Explain the concept of ethics as a branch of philosophy and its application in guiding moral decisions within scientific research.
- **Understand the ethical framework for psychological research:** Identify the four fundamental moral principles, which are weighing risks against benefits, acting responsibly and with integrity, seeking justice, and respecting people's rights and dignity, and describe how they apply to research participants, the scientific community, and society.
- **Evaluate risks and benefits in research:** Analyse how risks and benefits are assessed in psychological studies, including examples such as Milgram's obedience study, and discuss the challenges in balancing them.
- **Explain the importance of integrity in research:** Describe the role of honesty, transparency, and professional responsibility in building trust between researchers, participants, and the broader scientific community.
- **Discuss the principle of justice in research:** Examine the importance of fairness in distributing risks and benefits among participants and the consequences of historical injustices, such as the Tuskegee Syphilis Study.
- **Describe the concept of respect for participants' rights and dignity:** Explain the importance of informed consent, autonomy, and confidentiality in maintaining participants' rights and dignity.
- **Understand the ethical use of deception in research:** Describe the conditions under which deception can be ethically justified and the importance of thorough debriefing.
- **Develop effective informed consent and debriefing procedures:** Design clear and comprehensive informed consent and debriefing protocols to ensure participant understanding and well-being.
- **Navigate institutional approval processes:** Outline the role of Institutional Review Boards (IRBs) in evaluating and approving research proposals to ensure ethical compliance.

7.1. MORAL FOUNDATIONS OF ETHICAL RESEARCH

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Ethics is a branch of philosophy focused on understanding morality, including what it means to act morally and how individuals can achieve that standard. It also refers to a set of principles and guidelines that help people make moral decisions in specific fields, such as business, medicine, teaching, and scientific research.

In scientific research, especially studies involving human participants, ethical dilemmas can arise in many forms. To navigate these challenges, it is helpful to start with a general framework for understanding and addressing ethical issues.

A Framework for Understanding Research Ethics

Table 7.1.1 offers a clear framework for understanding the ethical considerations in psychological research. It focuses on four fundamental moral principles that guide ethical research practices: weighing risks against benefits, acting responsibly and with integrity, seeking justice, and respecting people's rights and dignity. These principles, adapted from the American Psychological Association (APA) Ethics Code, provide a foundation for making sound ethical decisions in research.

The table also outlines three key groups affected by scientific research. The first group is the research participants, who are directly involved in the study and may face risks or gain benefits from their participation. The second group is the scientific community, consisting of researchers, scholars, and professionals who depend on accurate and reliable research findings to advance knowledge. Finally, the third group is society as a whole, representing the broader public that can benefit from or be influenced by the outcomes of research studies. To ensure ethical integrity, researchers must consider how each moral principle applies to each of these groups. This framework encourages a balanced and thoughtful approach to addressing ethical concerns in research.

Table 7.1.1. A framework for thinking about ethical issues in scientific research

	Who is affected?		
Moral principle	**Research participants**	**Scientific community**	**Society**
Weighing risks against benefits			
Acting responsibly and with integrity			
Seeking justice			
Respecting people's rights and dignity			

Moral Principles in Research Ethics

Let us break down the key moral principles of research ethics and see how they apply to participants, the scientific community, and society.

Weighing Risks Against Benefits

Research ethics in psychology revolve around several core moral principles, which help ensure studies are conducted responsibly and ethically. One key principle is the importance of weighing risks against benefits. For research to be considered ethical, the potential benefits must outweigh any risks involved. Risks to research participants might include harm from ineffective or harmful treatments, physical or psychological distress from certain procedures, or breaches of privacy and confidentiality. On the other hand, participation can offer benefits such as access to helpful treatments, increased knowledge about psychology, the satisfaction of contributing to scientific progress, or even compensation in the form of money or academic credit.

The risks and benefits of research also extend beyond the participants themselves. For the scientific community, there is always a risk that valuable resources, such as time, funding, and effort, might be wasted on poorly designed or unproductive studies. Similarly, society at large faces risks when research findings are misunderstood or misapplied, leading to harmful consequences. A striking example of this was the flawed study falsely linking the MMR vaccine to autism, which caused widespread public health challenges. However, the benefits of well-conducted research are significant, advancing scientific knowledge and often resulting in meaningful improvements in health, education, and public policy.

Balancing these risks and benefits is not always straightforward because they do not always align. In many cases, participants might bear the majority of the risks, while the primary benefits are reaped by the scientific community or society as a whole. Stanley Milgram's 1963 study on obedience to authority serves as a powerful example of this ethical dilemma.

In Milgram's study, participants were told they were taking part in research on how punishment affects learning. They were instructed to administer electric shocks to another individual, who was actually a confederate pretending to be a participant. With every incorrect answer, the supposed shocks increased in intensity, and the confederate responded with recorded protests, complaints about heart problems, screams, and eventually silence. When participants hesitated or showed concern, the researcher would insist they continue. The results were both shocking and significant: most participants continued administering the shocks despite the apparent suffering they were causing. The study revealed nuanced insights into human obedience, with implications for understanding historical atrocities such as the Holocaust or events like the mistreatment of prisoners at Abu Ghraib.

However, the psychological toll on the participants was undeniable. Many experienced extreme distress, exhibiting symptoms such as sweating, trembling, stuttering, groaning, and nervous laughter. Some participants suffered uncontrollable seizures, and one individual's distress became so severe that the experiment had to be halted. Despite these outcomes, Milgram took significant steps to address the harm caused. He conducted thorough debriefing sessions, ensuring participants understood the true nature of the study and had an opportunity to recover emotionally. Most participants later reported feeling that their involvement was meaningful and expressed appreciation for contributing to important scientific knowledge.

The ethical debate surrounding Milgram's study continues to this day, raising the question of whether the knowledge gained was worth the emotional harm participants endured. It underscores the complexity of weighing risks against benefits in psychological research and serves as a reminder of the ongoing responsibility researchers have to carefully consider the ethical implications of their work.

Acting Responsibly and with Integrity

Researchers are expected to act responsibly and with integrity, ensuring their work is carried out with care, honesty, and professionalism. This means conducting research competently, fulfilling professional obligations, and being truthful in all aspects of their work. Integrity is essential because it fosters trust, which is the foundation of effective relationships, especially between researchers and participants. Participants must trust that researchers are being honest about the study's purpose, that promises such as maintaining confidentiality will be kept, and that every effort will be made to maximise benefits while minimising risks.

However, maintaining integrity is not always straightforward. In some cases, such as Milgram's obedience study, answering important research questions may require some level of deception. This creates an ethical conflict between advancing scientific knowledge for the greater good and being fully transparent with participants. Psychologists have developed strategies to address this conflict, which we will discuss shortly.

Trust must also extend beyond participants to include the scientific community and society as a whole. Researchers are responsible for conducting their studies thoroughly, competently, and honestly reporting their findings. When this trust is violated, the consequences can be severe. For example, the fraudulent study linking the MMR vaccine to autism misled both scientists and the public. Other researchers wasted valuable time and resources trying to replicate or address the flawed findings, while many parents avoided vaccinating their children. This misinformation ultimately led to outbreaks of preventable diseases like measles, mumps, and rubella, resulting in unnecessary suffering and even loss of life.

Seeking Justice

Researchers have a responsibility to ensure fairness in their work, treating participants equitably and distributing both the benefits and risks of research appropriately. Fair treatment includes providing participants with reasonable compensation for their time and effort and ensuring that no group bears an unfair share of the risks. For instance, in a study testing a promising new psychotherapy, one group might receive the therapy while another serves as a control group without treatment. If the therapy proves effective, justice would require offering the same treatment to the control group once the study concludes.

On a broader level, history reveals many examples where justice in research was ignored, particularly concerning vulnerable populations. Groups such as institutionalised individuals, people with disabilities, and racial or ethnic minorities have often faced disproportionate risks in scientific studies. One of the most infamous examples is the Tuskegee Syphilis Study, conducted by the U.S. Public Health Service between 1932 and 1972. In this study, poor African American men from Tuskegee, Alabama, were misled into believing they were receiving treatment for "bad blood". While they were given some basic medical care, they were intentionally left untreated for syphilis so researchers could observe the disease's progression. Even after penicillin became the standard treatment for syphilis in the 1940s, these men were still denied proper care and not given the option to leave the study. The experiment continued for decades until public outrage, sparked by investigative journalists and activists, brought it to an end. This tragic case serves as a powerful reminder of the importance of justice and fairness in research.

In 1997, 65 years after the study began and 25 years after it ended, President Bill Clinton issued a formal apology on behalf of the U.S. government. In his speech, he acknowledged the significant injustice faced by the men and their families:

> *So today America does remember the hundreds of men used in research without their knowledge and consent. We remember them and their family members. Men who were poor and African American, without resources and with few alternatives, they believed they had found hope when they were offered free medical care by the United States Public Health Service. They were betrayed.*

This apology stands as a solemn acknowledgment of the need for researchers to uphold justice, ensuring that every participant is treated fairly and that no group is unfairly burdened or exploited in the pursuit of scientific knowledge.

Respecting People's Rights and Dignity

Researchers have a responsibility to respect the rights and dignity of every participant. A key part of this is honouring participants' autonomy, which means recognising their right to make their own choices and take actions without being pressured or misled. Central to this principle is the concept of informed consent. This requires researchers to clearly explain the purpose, risks, and benefits of the study to participants and ensure they understand what their participation involves before agreeing to take part.

For informed consent to be meaningful, participants must have all the relevant information they need to make an informed decision. For example, in the Tuskegee Syphilis Study, participants were not told they had syphilis, nor were they informed that they would be intentionally denied treatment. If they had been given this critical information, it is unlikely they would have agreed to participate. Similarly, in Milgram's obedience study, participants were not warned about the severe emotional distress they might experience. If they had known they could be reduced to extreme anxiety and nervous breakdowns, many likely would have opted out. In both cases, the principle of informed consent was not properly upheld.

Another crucial aspect of respecting participants' rights and dignity is protecting their privacy. Participants have the right to decide what personal information they share and with whom. Researchers must safeguard participants' information through confidentiality, which means not sharing their personal data without consent or legal justification. Ideally, researchers should also aim for anonymity, where participants' names and any other identifiable information are not collected at all. This approach provides the highest level of privacy protection, ensuring participants can contribute to research without fear of their information being misused or exposed.

Unavoidable Ethical Conflict in Research

Ethical conflicts in psychological research are almost impossible to avoid. Since very few studies are completely risk-free, there is often a trade-off between potential risks and benefits. Research that benefits one group, such as the scientific community or society, can sometimes pose risks to another group, such as the research participants. Additionally, maintaining complete honesty with participants is not always possible, especially when deception is necessary to study certain behaviours accurately.

Some ethical dilemmas are relatively easy to resolve. For example, most people would agree that deceiving participants and causing them physical harm would not be justified simply to fill a minor gap in research knowledge. However, other ethical conflicts are far more complex, and even well-meaning, experienced researchers can disagree on how to address them.

A well-known example comes from a study on personal space conducted in a public men's restroom (Middlemist et al., 1976). The researchers wanted to see if the presence of another person nearby affected how long it took men to start urinating. To do this, they secretly observed participants without their consent. Critics argued that this study was an unjustified violation of human dignity (Koocher, 1977).

However, the researchers had carefully considered the ethical concerns and determined that the potential benefits outweighed the risks. They even interviewed preliminary participants and found that none were particularly bothered by the observation (Middlemist et al., 1977).

The key takeaway is that while ethical conflicts cannot always be eliminated, they can be managed responsibly. This involves carefully thinking through the ethical implications of a study, minimising risks, and balancing those risks against the potential benefits. Researchers must also be prepared to explain their ethical decisions, seek feedback from peers, and ultimately take responsibility for their actions.

References

Koocher, G. P. (1977). Bathroom behavior and human dignity. *Journal of Personality and Social Psychology, 35*(2), 120–121. https://doi.org/10.1037/0022-3514.35.2.120

Middlemist, R. D., Knowles, E. S., & Matter, C. F. (1976). Personal space invasions in the lavatory: Suggestive evidence for arousal. *Journal of Personality and Social Psychology, 33*(5), 541–546. https://doi.org/10.1037/0022-3514.33.5.541

Middlemist, R. D., Knowles, E. S., & Matter, C. F. (1977). What to do and what to report: A reply to Koocher. *Journal of Personality and Social Psychology, 35*(2), 122–124. https://doi.org/10.1037/0022-3514.35.2.122

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

7.2. FROM MORAL PRINCIPLES TO ETHICS CODES

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

The principles of weighing risks against benefits, acting with integrity, seeking justice, and respecting people's rights and dignity form a strong foundation for thinking about the ethics of psychological research. These principles are widely accepted and offer a common ground for ethical decision-making. However, even when people agree on these general principles, they may still disagree on how to handle specific ethical dilemmas that arise during research.

For this reason, detailed and enforceable ethics codes have been created to address recurring ethical issues and offer clear guidance for researchers. In this section, we will start with a brief look at the history of these ethics codes and then focus on the one most relevant to psychological research: the American Psychological Association (APA) Ethics Code.

Historical Overview

One of the first major ethics codes was the Nuremberg Code, created in 1947 following the trials of Nazi physicians who had conducted cruel and inhumane experiments on concentration camp prisoners during World War II. The code established ten key principles, emphasising the importance of carefully weighing risks against benefits and ensuring informed consent from participants. Many of the accused physicians were convicted and either imprisoned or sentenced to death based on these standards.

In 1964, the Declaration of Helsinki was introduced by the World Medical Association as an extension of the Nuremberg Code. This declaration added the requirement for researchers to create a written protocol, a detailed research plan, which must be reviewed by an independent ethics committee. The Declaration of Helsinki has undergone multiple revisions, with the most recent one occurring in 2004.

In the United States, growing concerns about unethical studies like the Tuskegee Syphilis Study led to the publication of the Belmont Report in 1978. This report highlighted three core principles:

1. **Justice:** Research must fairly distribute risks and benefits across different societal groups.

2. **Respect for Persons:** Researchers must respect individuals' autonomy and provide extra protection for those with diminished autonomy, such as children or prisoners. This principle supports the need for informed consent.
3. **Beneficence:** Researchers must aim to maximise benefits while minimising harm to participants and society.

The Belmont Report became the foundation for a set of laws known as the Federal Policy for the Protection of Human Subjects. These regulations require institutions receiving federal funding, such as universities and hospitals, to establish an Institutional Review Board (IRB). The IRB is responsible for reviewing research proposals to ensure they meet ethical standards. An IRB must include at least five members with diverse backgrounds, including scientists, non-scientists, men, women, and at least one person unaffiliated with the institution. The IRB evaluates research proposals to ensure that risks are minimised, benefits outweigh risks, participants are treated fairly, and informed consent is appropriately obtained.

Federal regulations classify research into three levels of risk:

- **Exempt Research:** This category involves minimal risk, such as studies on standard educational practices, surveys with non-sensitive topics where confidentiality is maintained, or research using publicly available data. Once approved, exempt research does not require ongoing IRB review.
- **Expedited Research:** This level includes research with slightly higher risk but still falls within the category of "minimal risk". Examples include certain psychological tests or studies involving standard physical or psychological assessments. Expedited reviews are conducted by either one IRB member or a small committee operating under the IRB's authority.
- **Full-Board Review Research:** Research that poses risks greater than minimal risk must undergo a full IRB review. In this process, the entire board evaluates the study to ensure all ethical standards are met.

These guidelines and review processes are essential for maintaining the balance between advancing scientific knowledge and protecting the rights and well-being of research participants.

Ethics Codes

For those interested in exploring the ethics codes discussed in this section, the Office of Human Subjects Research at the National Institutes of Health (NIH) provides access to the full texts. These documents are highly recommended reading, as they are generally brief, clear, and easy to understand, except for the Federal Policy, which is more detailed.

- The Nuremberg Code
- The Declaration of Helsinki
- The Belmont Report

- Federal Policy for the Protection of Human Subjects.

You can find these ethics codes on the Ethical Codes and Research Standards website.

APA Ethics Code

The APA's Ethical Principles of Psychologists and Code of Conduct, often referred to as the APA Ethics Code, was first introduced in 1953 and has been updated several times, most recently in 2010. This code outlines approximately 150 specific ethical standards that psychologists and their students must follow. While many of these standards focus on clinical practices, such as advertising services, managing fees, and maintaining professional boundaries, Standard 8: Research and Publication is particularly relevant for research ethics. Table 7.2.1 lists and simplifies the key aspects of APA Ethics Code Standard 8: Research and Publication.

Table 7.2.1. Key aspects of APA Ethics Code Standard 8: Research and Publication

Institutional Approval:	Before starting any research, psychologists must submit accurate research proposals and gain approval from an Institutional Review Board (IRB). Research must then follow the approved protocol.
Informed Consent:	Researchers must ensure participants understand the study before agreeing to participate. This includes explaining the study's purpose, duration, procedures, potential risks, benefits, and confidentiality measures. Participants must also know they can withdraw at any time without consequence.
Recording Voices and Images:	Researchers must obtain consent before recording participants' voices or images unless the study involves public, natural observations without any risk of harm or identification.
Research with Vulnerable Participants:	When working with clients, patients, students, or subordinates, psychologists must ensure participation is voluntary and free from pressure or negative consequences for declining or withdrawing.
When Informed Consent Is Not Required:	In some low-risk studies, such as anonymous surveys or archival research, psychologists may be exempt from obtaining informed consent, provided participants are not at risk of harm or privacy breaches.
Inducements for Participation:	Researchers should avoid offering excessive rewards or incentives that might pressure individuals into participating.
Deception in Research:	Deception is only allowed if it is essential for the study's purpose, poses no significant harm, and participants are thoroughly debriefed afterwards.
Debriefing:	After participation, psychologists must provide participants with complete information about the study, clarify any misconceptions, and address any harm caused by the research process.
Use of Animals in Research:	When using animals in research, psychologists must follow federal, state, and professional guidelines to ensure humane care, minimise harm, and properly train all personnel involved.
Reporting Results:	Researchers must not fabricate or falsify data. If errors are discovered after publication, they must take steps to correct them.
Plagiarism:	Psychologists must not present someone else's work or data as their own, even if citations are included.
Publication Credit:	Authorship should reflect the actual contributions made by each person involved in the research. Faculty advisors should ensure students receive appropriate credit, especially when the research is based on a dissertation.
Duplicate Publication:	Data that have already been published should not be presented as new, original findings unless properly acknowledged.
Sharing Research Data:	After publication, researchers must share their data with other professionals who wish to verify findings, provided participant confidentiality is protected.
Peer Review Responsibilities:	Psychologists reviewing materials for publication or funding must respect confidentiality and avoid misusing privileged information.

A more detailed version of the full ethics code is available on the APA Ethics Code website.

Informed Consent

Informed consent is about ensuring participants understand and agree to take part in a study after being fully informed about what it involves. This includes explaining the study's purpose, procedures, potential risks and benefits, their right to refuse participation or withdraw at any time, and any legal limitations on confidentiality. For example, in some states, researchers are legally required to report evidence of child abuse or other crimes.

While informed consent often involves participants reading and signing a consent form, the form itself is not enough. Many participants either skim through the form or fail to understand its content. Some mistakenly believe that signing the form means they are giving up their right to take legal action (Mann, 1994). Therefore, it is good practice for researchers to go beyond the form. This means explaining the study verbally, answering questions, demonstrating procedures when appropriate, and reminding participants of their right to withdraw at any time.

There are also situations where informed consent is not required. If the study poses no risk and involves everyday activities, formal consent might not be necessary. For instance, observing whether people hold doors open in public spaces does not require consent. Similarly, if a college instructor compares two standard teaching methods across different class sections, informed consent would not be needed because both methods fall within ordinary educational practices.

Deception in Psychological Research

In psychological research, deception can take many forms. Researchers might mislead participants about a study's purpose, use actors (called confederates), employ fake equipment like Milgram's shock generator, or give false feedback (e.g., telling someone they performed poorly on a test when they actually did well). Deception can also involve leaving out key details about the study's true purpose, even if no outright lies are told. For example, in a study on incidental learning, participants might assume they will be tested on memorising words from a list. However, the real test might focus on something unexpected, like their memory of the room's layout or the research assistant's appearance.

Some researchers believe deception is rarely, if ever, ethically acceptable. They argue that it undermines informed consent, disrespects participants' dignity, risks causing distress, reduces trust in researchers, and potentially harms the credibility of the entire field (Baumrind, 1985).

However, the APA Ethics Code takes a more balanced stance. It allows deception if four key conditions are met:

1. The study's benefits outweigh the risks.
2. Participants are unlikely to suffer harm.
3. The research question cannot be answered without using deception.

4. Participants are informed about the deception as soon as possible, typically during debriefing.

This approach recognises that not all deception is equally harmful. For example, Milgram's famous study caused severe psychological stress through significant deception. In contrast, a simple incidental learning study where participants are mildly misled about a memory test poses far less risk.

Additionally, some important research questions simply cannot be answered without using deception. If participants know in advance that a study focuses on obedience, aggression, or helping behaviour, their awareness could influence their actions. As a result, the study's findings might no longer reflect real-world behaviour.

Debriefing in Psychological Research

Debriefing, outlined in Standard 8.08 of the APA Ethics Code, is the process of explaining a study's purpose to participants after their involvement ends. This step is especially important when deception was used. Researchers must clarify the true goals of the study, reveal any misleading information, and correct any misunderstandings participants might have.

Debriefing also focuses on minimising any harm or discomfort caused by the study. For example, in an experiment designed to study how sadness affects memory, participants might have been put in a sad mood by thinking about unhappy memories, watching a sad video, or listening to melancholy music. During debriefing, researchers would actively help participants return to a neutral or positive emotional state, perhaps by showing an uplifting video, playing cheerful music, or encouraging them to focus on happy thoughts.

Research with Nonhuman Animal Subjects

Standard 8.09 of the APA Ethics Code focuses on the humane treatment and care of nonhuman animal subjects in psychological research. While most psychological studies today do not involve animals, they still play an essential role in certain areas, such as understanding learning and behaviour, exploring brain functions, and developing treatments for psychological disorders.

The use of animals in research has sparked significant ethical debate. Critics argue that animals cannot provide informed consent and may be subjected to distressing conditions, such as confinement, food or water deprivation, painful procedures, surgeries, or even euthanasia. However, some research is far less invasive, involving simple observation in natural or controlled environments.

Supporters of animal research highlight its significant contributions to both human and animal well-being. Animal studies have led to breakthroughs in behavioural therapies, pain management techniques, and medications for mental health disorders. Additionally, these studies have benefited animals themselves, offering more humane methods for managing animal populations compared to practices like poisoning or shooting.

The APA takes a balanced stance, allowing research with nonhuman animals when the potential benefits outweigh the risks. Researchers are required to use alternative methods whenever possible. If animal subjects must be used, they must be housed, fed, and cared for in humane conditions, and any harm must be minimised.

For more details on the APA's guidelines for animal research, you can visit the APA Committee on Animal Research and Ethics website.

Scholarly Integrity

Standards 8.10 to 8.15 of the APA Ethics Code focus on maintaining honesty and transparency in research and publishing. At the core of these guidelines are key principles:

Researchers must never fabricate or falsify data and must not plagiarise. Plagiarism involves using someone else's words or ideas without giving proper credit. Proper acknowledgment means using quotation marks for direct quotes and including a clear citation for any borrowed ideas or phrasing. Additionally, self-plagiarism is unethical. This happens when researchers recycle their own previously published material and present it as new work, just as students should not submit the same paper for multiple classes.

Other important aspects of scholarly integrity include avoiding duplicate publication, where the same dataset is published twice as if it were new. Researchers are also expected to share their data with other qualified researchers for verification and further analysis, provided confidentiality and participant privacy are protected.

When acting as peer reviewers, researchers must respect the confidentiality of the unpublished research they review. They should not use or disclose any information from these manuscripts.

Lastly, authorship credit must accurately reflect each person's contribution to the research. The order of authors' names should be based on the significance of their contributions. It is unethical to list someone as an author who only made a minor contribution, such as running an analysis, or for a faculty member to claim first authorship on a project primarily conducted by a student.

References

Baumrind, D. (1985). Research using intentional deception: Ethical issues revisited. *American Psychologist, 40*(2), 165–174. https://doi.org/10.1037/0003-066X.40.2.165

Mann, T. (1994). Informed consent for psychological research: Do subjects comprehend consent forms and understand their legal rights? *Psychological Science, 5*(3), 140-143. https://doi.org/10.1111/j.1467-9280.1994.tb00650.x

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

7.3. PUTTING ETHICS INTO PRACTICE

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

This section offers practical guidance for conducting ethical research in psychology. It is essential to recognise that ethical considerations are not limited to the data collection phase but emerge at every stage of the research process, from initial planning and study design to publication and the broader dissemination of findings.

Understand and Embrace Your Ethical Responsibilities

The American Psychological Association (APA) Ethics Code states clearly that "Lack of awareness or misunderstanding of an ethical standard is not itself a defense to a charge of unethical conduct." This means that as a researcher, it is your responsibility to fully understand and uphold ethical standards throughout your work.

To begin, make sure you are familiar with the APA Ethics Code, particularly the sections relevant to research. You should be able to distinguish between minimal risk research and at-risk research and understand your institution's specific policies and procedures for ethical approval. This includes knowing how to properly prepare and submit your research proposal to the Institutional Review Board (IRB) for review.

If your research is part of a course requirement, additional course-specific policies and expectations might apply. If you encounter an ethical question or are unsure about any standard, policy, or procedure, seek clarification immediately. You can do this by:

- Referring to the relevant ethics codes
- Researching how similar ethical issues have been addressed by others
- Consulting experienced researchers, your IRB, or your course instructor.

Ultimately, the ethical responsibility for your research rests with you. Taking the time to understand

and follow ethical guidelines not only protects your participants but also maintains the integrity of your research and contributes to the credibility of the scientific community.

Identifying and Minimising Risks in Research

When designing your study, it is essential to identify and minimise potential risks to participants. Risks can include physical harm, psychological distress, or breaches of confidentiality.

Start by carefully listing all possible risks. Keep in mind that researchers often underestimate the seriousness of risks or overlook them entirely. For example, a student researcher testing people's sensitivity to violent images planned to show participants graphic photos from crime and accident scenes. Because she was an emergency medical technician (EMT), she was desensitised to such images and failed to realise how deeply disturbing they might be for others.

Also, remember that certain risks may affect only specific participants. For example, most people might have no issue answering a survey about their fear of crime, but someone who has been a victim of a violent crime might find those questions upsetting. To avoid such oversights, seek input from others, including collaborators, experienced researchers, and even non-researchers who can offer a participant's perspective.

Once risks are identified, you can often reduce or eliminate them in three main ways:

1. Modify the Research Design

Simplify or shorten procedures to minimise frustration and fatigue. If your study uses upsetting materials, consider using less distressing alternatives. For example, instead of graphic accident scene photos, use milder images similar to those shown in newspapers.

A good example of this approach comes from Jerry Burger's (2009) replication of Milgram's obedience study. Burger stopped participants from administering shocks beyond 150 volts, knowing that Milgram's original participants experienced the most severe stress after this point. By doing this, Burger could still compare his findings with Milgram's results while avoiding unnecessary psychological harm to participants. Interestingly, Burger found that modern participants were just as obedient as those in Milgram's original study.

2. Pre-Screen Participants

Use a pre-screening process to identify participants who may be at high risk of harm. The informed consent process can also help by warning participants about sensitive topics and reminding them they can withdraw at any time.

Pre-screening might involve administering surveys or interviews to identify physical or psychological conditions that might make participation risky. For example, Burger's study included extensive pre-screening with questionnaires and a clinical psychologist's interview to eliminate high-risk participants before they took part in the study.

3. Protect Confidentiality

Maintaining confidentiality is crucial to protecting participants. Store signed consent forms separately from research data to ensure participants' identities remain anonymous.

Additionally:

- Only collect essential personal information required for your research question. If details like sexual orientation or ethnicity are not directly relevant, do not ask for them.
- Be cautious with data collection environments. For example, conducting oral surveys in public places like shopping malls or distributing questionnaires in shared classrooms can lead to unintentional breaches of confidentiality. Instead, administer surveys in private or use strategies to ensure responses remain secure.

By carefully identifying risks, refining your study design, pre-screening participants, and protecting confidentiality, you can create a safer and more ethical research experience for everyone involved.

Reducing and Justifying Deception in Research

Deception in research can take many forms, and it does not always involve directly misleading participants. It can also include allowing participants to make incorrect assumptions about the study or withholding key details about its purpose or design. To ensure ethical practices, it is important to identify and minimise all forms of deception in your study.

Is Deception Necessary?

According to the APA Ethics Code, deception is only ethically acceptable if there is no other way to answer your research question. If your study includes any type of deception, ask yourself whether it is truly essential.

For example, imagine you want to study whether the age of college professors affects students' expectations about their teaching ability. You might plan to show participants photos of people and ask them to rate their teaching ability, claiming the images are of real college professors. However, if the photos are actually of your friends and family, this would constitute deception. Instead, you could simply tell participants that the photos represent college professors and ask them to rate them as if they were. This approach removes the need for deception while still allowing you to answer your research question effectively.

Balancing Informed Consent and Research Validity

It is generally acceptable to withhold your specific research question until the debriefing stage, as long as you fully inform participants about the procedures, risks, and benefits during the informed consent process.

In the example about age and teaching expectations, you would not need to tell participants that you are studying how age affects their perceptions. Sharing this detail upfront might change their behaviour, where some might unconsciously rate older and younger "professors" differently because they think that is what you expect, while others might overcorrect to avoid appearing biased.

To address this, you can include a brief clarification during the consent process. You might explain to participants, either orally, in writing, or both, that while the procedures, risks, and benefits have been fully described, you will wait until after the study to reveal the exact research question. In essence, participants are giving their consent to be temporarily misled or to have certain information withheld until debriefing.

Balancing Risks and Benefits in Research

After identifying and minimising risks in your research, the next step is to weigh those risks against the potential benefits. This process involves considering all possible benefits, not just for participants, but also for science and society as a whole.

If you are a student researcher, do not forget that one of the benefits is the experience and knowledge you will gain about conducting psychological research. These skills will help you succeed in your studies, graduate school, or your future career.

Minimal vs. Greater Than Minimal Risk

If your research involves minimal risk, where the risks are no greater than what people encounter in daily life or during routine medical or psychological exams, then even a small benefit can justify the study.

However, if your research involves more than minimal risk, the benefits must be more substantial. For example:

- If your study might cause participants emotional distress or discomfort, it must address a meaningful scientific question or have clear practical value.
- It would be unethical to subject participants to pain, fear, or embarrassment without a valid scientific purpose or simply to satisfy personal curiosity.

Ethical Guidelines for Higher Risk Research

In general, psychological research that has the potential to cause serious or long-lasting harm is rarely justified unless the benefits are exceptionally significant.

When weighing risks and benefits, always ask:

- Is the research question important and valuable?
- Are the risks minimised as much as possible?
- Do the benefits outweigh any potential harm?

By carefully balancing risks and benefits, you ensure that your research is not only scientifically valuable but also ethically responsible.

Developing Informed Consent and Debriefing Procedures

After finalising your research design, the next step is to establish clear informed consent and debriefing procedures.

Informed Consent

Start by determining whether informed consent is required under APA Standard 8.05. If it is, follow these key steps:

1. **Provide Clear Information During Recruitment:** Whether you are recruiting participants through word of mouth, flyers, or an online participant pool, share as much information as possible about the study upfront. This allows people who might find the study uncomfortable or objectionable to opt out.
2. **Prepare a Script or Talking Points:** Write a clear and simple explanation of your study in everyday language. Include details about the procedure, potential risks and benefits, and participants' right to withdraw at any time.
3. **Create an Informed Consent Form:** Develop a form that covers all key elements outlined in Standard 8.02a. Participants should read and sign this form after you have explained the study to them. Many institutions or instructors provide sample consent forms you can customise. If not, you can find reliable templates online.
4. **Address Deception (if applicable):** If your study involves withholding certain information or using deception, state clearly (both orally and in writing) that some details will be revealed during debriefing.

Debriefing

Debriefing is just as important as informed consent and follows a similar approach.

1. **Use a Script or Talking Points:** Do not rely solely on written debriefing forms. Instead, prepare a clear explanation in simple language.
2. **Explain the Research Fully:** Share the study's true purpose and design, including what happened in conditions participants were not exposed to.
3. **Address Deception Honestly:** If deception was used, reveal it as soon as possible, apologise, and explain why it was necessary. Correct any misunderstandings participants might have as a result.
4. **Offer Resources and Support:** Debriefing is an opportunity to provide practical resources or referrals that might benefit participants. For example, in a study about attitudes toward domestic abuse, you could offer pamphlets and contact information for counselling services.

Plan Adequate Time

Both informed consent and debriefing require time. Rushing through either process can compromise their effectiveness and leave participants feeling uninformed or undervalued.

Securing Institutional Approval

Before starting your study, you will need to obtain institutional approval based on your institution's or course's specific policies and procedures.

This process typically involves writing a detailed research protocol that includes:

- the purpose of your study
- the research design and procedure you will follow
- a clear explanation of risks and benefits for participants
- steps taken to minimise risks and protect participants
- your informed consent and debriefing procedures.

While the approval process might seem like just another hurdle, it is actually a valuable opportunity to carefully think through the ethical aspects of your research. It also allows you to consult with experienced reviewers who can provide useful insights and fresh perspectives.

If the Institutional Review Board (IRB) raises questions or suggests changes, respond promptly and thoughtfully. This might involve making adjustments to your research design or procedures and resubmitting your protocol for further review. Approaching this step with an open mind will not only help you meet ethical requirements but also improve the overall quality and integrity of your study.

Staying Ethical Throughout the Research Process

Ethical responsibility does not stop once your study receives institutional approval. It is essential to follow the approved protocol carefully and seek additional approval if you need to make any significant changes.

During the research process, pay close attention to participants' reactions and remain alert for any unexpected responses or signs of distress. Gather feedback during debriefing to identify any concerns participants might have. For example, one criticism of Milgram's obedience study was that although he could not have predicted participants' severe stress reactions, he should have adjusted the procedure after observing the first few participants' experiences.

You must also protect confidentiality throughout the study. Keep consent forms and participant data separate and secure to prevent accidental or intentional breaches of privacy. Ensure that no one outside the research team has access to participants' personal information.

Your commitment to integrity extends to publication and beyond. Clarify authorship roles early with your collaborators, making sure credit accurately reflects contributions. Avoid plagiarism by properly citing sources and never reusing your previously published work without acknowledgment.

Most importantly, stay honest about your findings. Your role as a scientist is to report your results truthfully, even if they do not align with your predictions. Unexpected outcomes often lead to valuable new insights and can be just as important, if not more so, than anticipated ones.

References

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist, 64*(1), 1–11. https://doi.org/10.1037/a0010932

Chapter Attribution

Content adapted, with editorial changes, from:

7.4. SUMMARY

By Marc Chao

Summary

Ethics in psychological research revolves around principles that safeguard the rights and well-being of participants, ensure the integrity of scientific inquiry, and benefit society. Adapted from the APA Ethics Code, these principles, which include weighing risks against benefits, acting responsibly and with integrity, seeking justice, and respecting participants' rights and dignity, provide a framework for ethical decision-making. Researchers must balance the potential benefits of their studies against risks such as physical or psychological harm, while maintaining transparency and fairness in their methods. Justice requires equitable treatment of participants and fair distribution of risks and rewards, underscoring the importance of avoiding historical injustices like the Tuskegee Syphilis Study. Respecting participants' rights and dignity includes obtaining informed consent, safeguarding confidentiality, and protecting autonomy, ensuring that research practices uphold human values and trust.

To address recurring ethical dilemmas, detailed ethics codes such as the APA Ethics Code have been developed, drawing on foundational guidelines like the Nuremberg Code, the Declaration of Helsinki, and the Belmont Report. These frameworks establish standards for minimising risks, upholding informed consent, and ensuring equitable treatment of vulnerable populations. The APA Ethics Code further outlines specific requirements, including obtaining institutional approval, maintaining transparency, and responsibly managing ethical challenges like deception or animal research. It also emphasises scholarly integrity by prohibiting data fabrication, plagiarism, and duplicate publication while promoting transparency, proper authorship credit, and data sharing. Together, these guidelines ensure that research advances scientific knowledge responsibly while prioritising the welfare of participants and the broader community.

Ethical research in psychology requires vigilance at every stage, from planning and design to dissemination and publication. Researchers must adhere to the APA Ethics Code by obtaining

institutional approval, carefully identifying and minimising risks, and ensuring informed consent and effective debriefing. Safeguarding participants involves refining research designs, pre-screening for vulnerabilities, and protecting confidentiality. When deception is necessary, it must be justified, minimised, and addressed transparently during debriefing. Weighing risks against benefits ensures that studies are scientifically valuable and ethically defensible, particularly when involving higher risks. By engaging in thoughtful risk mitigation, maintaining scholarly integrity, and following approved protocols, researchers uphold ethical standards, enhance the credibility of their findings, and contribute meaningfully to the advancement of psychological science.

CHAPTER 8: PSYCHOLOGICAL MEASURES

Measurement is a fundamental aspect of psychological research, serving as the bridge between abstract theoretical concepts and observable, quantifiable data. In psychology, measurement extends beyond physical attributes like height and weight to encompass complex mental states, emotions, and cognitive abilities. This chapter explores how psychologists systematically assign scores to represent intangible characteristics such as memory, self-esteem, or anxiety, ensuring these measurements are both meaningful and reliable.

At its core, psychological measurement relies on constructs, which are abstract concepts like intelligence, fear, or extraversion that cannot be directly observed but can be inferred through patterns of behaviour, self-reported experiences, or physiological responses. To bring these constructs to life in research, psychologists use operational definitions, which specify how a construct will be measured, whether through surveys, observations, or biological markers.

The chapter also delves into the levels of measurement, ranging from simple categorisations (nominal) to highly precise scales with meaningful zero points (ratio). Understanding these levels helps researchers choose appropriate statistical techniques for analysing their data. Furthermore, the chapter emphasises the importance of reliability, which is the consistency of a measurement tool, and validity, which is the degree to which it measures what it claims to measure. Both are essential for drawing accurate conclusions from research findings.

By examining key concepts such as operational definitions, measurement levels, reliability, and validity, this chapter provides a comprehensive foundation for understanding how psychological traits and processes are systematically assessed. Readers will gain insight into the rigorous standards psychologists adhere to when transforming abstract ideas into measurable outcomes, ensuring the credibility and integrity of psychological research.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand the concept of psychological measurement:** Explain what psychological measurement is and how abstract psychological traits can be systematically measured.

- **Differentiate between psychological constructs and observable traits:** Distinguish between directly measurable variables (e.g., height, weight) and abstract constructs (e.g., self-esteem, fear), and explain why constructs require operational definitions.
- **Explain the importance of conceptual and operational definitions:** Discuss the significance of defining constructs both conceptually and operationally to ensure valid and reliable measurement.
- **Identify and differentiate measurement methods:** Describe self-report, behavioural, and physiological measures and provide examples of how each method is used in psychological research.
- **Evaluate reliability and validity in measurement:** Define reliability and validity, explain their importance in psychological measurement, and describe how they ensure consistent and accurate results.
- **Understand the principle of converging operations:** Explain how using multiple operational definitions strengthens the validity of measuring psychological constructs.

8.1. UNDERSTANDING PSYCHOLOGICAL MEASUREMENT

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

What Is Measurement?

Measurement involves assigning scores to individuals in a way that those scores represent specific characteristics or traits of those individuals. This concept applies to everyday situations, such as stepping on a bathroom scale to measure weight or using a meat thermometer to check the internal temperature of a turkey. It is also central to scientific disciplines. In physics, for example, measuring an object's potential energy involves determining its mass and height, then using a formula that includes Earth's gravitational acceleration (9.8 m/s^2) to calculate the final value. The resulting number represents the object's potential energy.

This same principle applies to psychological measurement, also known as psychometrics. In psychology, the goal is to systematically assign scores to represent intangible traits or mental states. For instance, a cognitive psychologist interested in working memory capacity might use a backward digit span task. In this task, the psychologist reads a series of digits and asks the participant to repeat them in reverse order. The length of the longest digit sequence correctly repeated serves as the participant's score, representing their working memory capacity.

Similarly, a clinical psychologist might want to measure depression levels. To do this, they could use the Beck Depression Inventory (BDI), a 21-item questionnaire where participants rate how frequently they have experienced symptoms like sadness or fatigue over the past two weeks. The total score from these ratings reflects the participant's current level of depression.

The key takeaway is that measurement does not require a specific tool or instrument. Instead, it relies on a systematic method for assigning scores in a way that accurately reflects the characteristic being measured. Whether it is working memory, depression, or physical weight, the measurement process must follow a structured approach to ensure consistency and meaningful results.

Psychological Constructs

Some variables studied in psychology, such as age, height, weight, and birth order, are relatively easy to measure. For example, asking someone their age usually provides an accurate answer, and if someone is unsure or unwilling to share their weight, a bathroom scale offers an objective measurement. However, most psychological variables are not so simple to measure. You cannot determine someone's intelligence just by looking at them, nor can you measure self-esteem with a scale. These more abstract variables are called constructs and include traits like extraversion, emotional states like fear, attitudes such as opinions on taxes, and abilities like athleticism.

Psychological constructs cannot be observed directly for a couple of reasons. First, they often represent tendencies to think, feel, or act in certain ways rather than observable actions at any given moment. For instance, saying that a student is highly extraverted does not mean she's always outgoing. At this moment, she might be sitting quietly reading a book. Instead, extraversion reflects her general tendency to be outgoing and socially engaged across various situations.

Second, constructs often involve internal processes that are not visible to an observer. For example, fear activates certain parts of the nervous system, triggers specific thoughts and feelings, and may lead to behaviours like avoiding danger, all of which may not be apparent to someone watching. Importantly, constructs like extraversion or fear are not reduced to one specific behaviour, thought, or biological response. Instead, each construct acts as a summary of a broader pattern of behaviours and internal processes.

A conceptual definition of a construct explains the behaviours and internal processes that make up that construct and outlines how it relates to other variables. For example, neuroticism is defined as a tendency to experience negative emotions like anxiety, anger, and sadness across different situations. This definition might also mention that neuroticism has a genetic basis, remains relatively stable over time, and is associated with a higher tendency to experience physical pain and other symptoms.

Students sometimes wonder why researchers do not simply rely on dictionary definitions for constructs like self-esteem or neuroticism. The reason is that scientific constructs often have no direct counterpart in everyday language. For example, working memory capacity is not a term you would typically find in casual conversation. More importantly, scientific definitions are far more detailed and precise than dictionary definitions. Researchers aim to create definitions that accurately reflect reality and are refined through empirical testing and adjustment based on evidence.

In psychology, it is common to find multiple definitions for the same construct in the research literature. This happens because researchers are continually testing, refining, and sometimes replacing older definitions with ones that better explain their findings. In some cases, there is an ongoing debate about which definition is most accurate. This iterative process is central to the scientific study of psychological constructs and helps ensure that these abstract ideas are measured and understood as precisely as possible.

Operational Definitions

An operational definition explains a variable in terms of how it is specifically measured in a study. Psychologists typically measure variables in three main ways: self-report measures, behavioural measures, and physiological measures.

In self-report measures, participants describe their own thoughts, feelings, or behaviours. For example, the Rosenberg Self-Esteem Scale asks people to rate statements about their self-worth.

In behavioural measures, researchers observe and record actions or behaviours. These observations can happen in controlled laboratory settings or in natural environments. For example, working memory capacity can be measured using a backward digit span task, where participants repeat numbers in reverse order. In a more natural setting, Albert Bandura and his colleagues measured physical aggression by observing children play with a Bobo doll. They counted specific aggressive behaviours, such as hitting, kicking, or punching the doll. The number of these actions within a set time frame served as the operational definition of aggression in their study.

In physiological measures, researchers record biological processes such as heart rate, blood pressure, stress hormone levels, or brain activity. These measures provide objective data on participants' physical responses to stimuli or conditions.

For any single construct, there are often multiple valid operational definitions. Stress is a good example. Conceptually, stress can be defined as an adaptive response to a perceived threat, involving physiological, emotional, and behavioural changes. However, stress has been measured in many ways:

- The Social Readjustment Rating Scale evaluates stress by assigning points to life events, such as divorce or job change, based on their severity.
- The Hassles and Uplifts Scale focuses on everyday stressors like misplacing items or worrying about weight.
- The Perceived Stress Scale asks participants how frequently they feel nervous or overwhelmed.
- Physiological measures, such as blood pressure or cortisol levels, provide biological markers of stress.

When psychologists use multiple operational definitions for the same construct, either in one study or across different studies, they are applying the principle of converging operations. This approach assumes that different measures of the same construct should produce similar results.

For example, if different stress measures (e.g., self-report questionnaires and physiological indicators) correlate with each other and show consistent patterns, this strengthens confidence that the construct is being accurately measured. Studies have shown that various measures of stress all correlate with immune system functioning, reinforcing the conclusion that stress negatively affects immune health (Segerstrom & Miller, 2004).

Levels of Measurement

Psychologist S.S. Stevens introduced the idea that measurements can be categorised based on how much quantitative information they communicate about a variable. For example, in a 100-metre race, runners' performance can be recorded in two ways: simply by their rank order (1st, 2nd, 3rd) or by using a stopwatch to record exact times (11.5 seconds, 12.1 seconds). While both methods measure performance, the stopwatch provides more detailed information because it shows not only the order but also how much faster or slower one runner was compared to another.

As shown in Table 8.1.1, Stevens identified four levels of measurement, nominal, ordinal, interval, and ratio, each offering a different level of detail and determining which statistical methods are appropriate.

Nominal Level: Categorising Data

At the nominal level, data is grouped into categories or labels without any implied order. For example, asking participants about their marital status (single, married, divorced) or ethnicity involves nominal-level measurement. These labels indicate differences but do not suggest any ranking or order, as being "single" is not inherently higher or lower than being "married".

Key takeaway: Nominal scales classify data but do not rank it.

Ordinal Level: Ranking Data

At the ordinal level, data is ranked or ordered, but the intervals between ranks are not necessarily equal. For example, if people rate their satisfaction with a product as "very dissatisfied", "somewhat dissatisfied", "somewhat satisfied", or "very satisfied", the categories are ranked. "Very satisfied" is clearly higher than "somewhat satisfied", but the difference between these two categories might not be the same as the difference between "somewhat dissatisfied" and "very dissatisfied".

Similarly, in a race, the difference in time between the 1st and 2nd place finishers might be tiny, while the difference between 2nd and 3rd place could be much larger. Ordinal scales tell us who is higher or lower, but not how much higher or lower.

Key takeaway: Ordinal scales rank data, but intervals between ranks may not be consistent.

Interval Level: Equal Intervals, No True Zero

The interval level provides more information by ensuring that the differences between values are consistent across the scale. A good example is the Celsius or Fahrenheit temperature scales. The difference between 30°C and 40°C is the same as the difference between 80°C and 90°C.

However, interval scales lack a true zero point. For instance, 0°C does not mean the absence of temperature; it is just another point on the scale. This means ratios do not hold meaningful comparisons because the zero point is arbitrary, so you cannot say 80°C is "twice as hot" as 40°C.

In psychology, IQ scores are considered interval-level measurements. A score of 0 does not mean no intelligence, and an IQ of 140 is not "twice as intelligent" as an IQ of 70. However, the difference between an IQ of 80 and 100 is the same as the difference between 120 and 140.

Key takeaway: Interval scales have equal intervals but no true zero, making ratio comparisons meaningless.

Ratio Level: True Zero Point

The ratio level is the most precise level of measurement because it has equal intervals and a true zero point, indicating the absence of the characteristic being measured. Examples include weight (in kilograms), height (in metres), and income (in dollars).

For instance, someone who weighs 0 kg truly has no weight, and someone with $50 has exactly twice as much money as someone with $25. The Kelvin temperature scale is another example because 0 K represents absolute zero, which is the complete absence of molecular motion.

Key takeaway: Ratio scales allow for meaningful comparisons of both intervals and ratios.

Table 8.1.1. Summary of levels of measurements

Level of Measurement	Category labels	Rank order	Equal intervals	True zero
NOMINAL	X			
ORDINAL	X	X		
INTERVAL	X	X	X	
RATIO	X	X	X	X

Reliability and Validity of Measurement

Measurement in psychology involves assigning scores to individuals to represent certain characteristics or traits accurately. However, when dealing with abstract constructs such as intelligence, self-esteem, or depression, researchers must ensure that these scores genuinely reflect the intended characteristic. To achieve this, psychologists conduct studies to confirm that their measurement tools function as expected. If the results suggest the measure is unreliable or invalid, it is either revised or abandoned altogether.

Imagine you have been dieting for a month. Your clothes fit more loosely, and friends have noticed your weight loss. If your bathroom scale shows you have lost 10 pounds, it aligns with your observations, and you would trust the scale. However, if it indicates a gain of 10 pounds, you would suspect it is broken and either fix or replace it. This analogy highlights how psychologists approach evaluating their measurement tools. Two key dimensions guide this evaluation: reliability and validity.

References

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychological Bulletin, 130*(4), 601–630. https://doi.org/10.1037/0033-2909.130.4.601

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

8.2. RELIABILITY

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Reliability refers to the consistency of a measurement tool. A reliable measure produces stable and consistent results across time, across different items within the same test, and across different observers. In psychology, reliability is assessed in three main ways: test-retest reliability, internal consistency, and inter-rater reliability.

Test-Retest Reliability

When researchers measure something they expect to remain stable over time, the results should also remain consistent. Test-retest reliability refers to how well a measurement tool produces similar results when used on the same people at different points in time.

For example, intelligence is generally considered a stable trait. If someone scores high on an intelligence test today, they should score similarly next week. If the test produces very different scores each time, it is not a reliable tool for measuring a stable construct like intelligence.

To evaluate test-retest reliability, researchers measure the same group of people twice using the same tool, usually with a gap of a few days or weeks between the two measurements. They then compare the two sets of scores using a correlation coefficient, which indicates how closely the two measurements match. A scatterplot is often used to visually represent this relationship.

For instance, in Figure 8.2.1, we can see that if university students take the Rosenberg Self-Esteem Scale twice, one week apart, and the two sets of scores have a correlation coefficient of +0.95, this indicates excellent reliability. In general, a correlation of +0.80 or higher suggests good test-retest reliability.

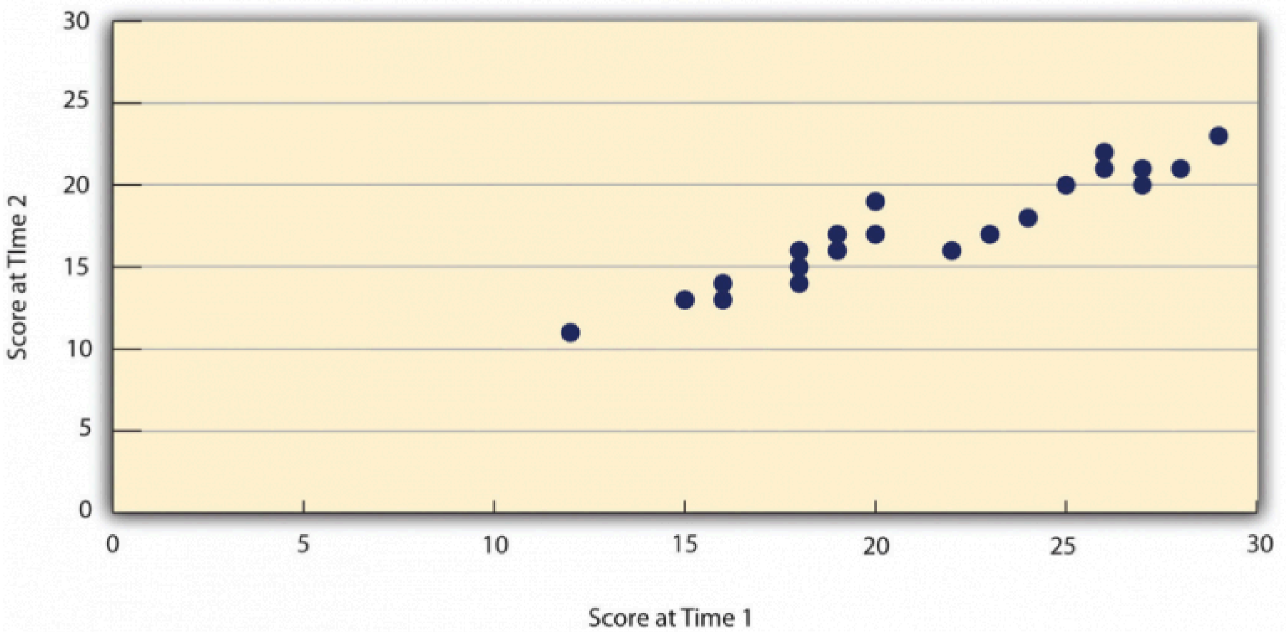


Figure 8.2.1. Test-retest correlation between two sets of scores of several college students on the Rosenberg Self-Esteem Scale, given two times a week apart by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

However, not all psychological constructs are expected to remain stable over time. Traits like intelligence, self-esteem, and personality dimensions (e.g., the Big Five traits) are typically consistent, so high test-retest reliability is expected. On the other hand, constructs like mood are naturally more variable. If a mood-measuring tool shows low test-retest reliability over a month, that is not necessarily a problem because mood is expected to change frequently.

Internal Consistency

Internal consistency refers to how well the items on a multi-item measure align with each other in assessing the same underlying construct. When researchers use multiple questions or tasks to measure something like self-esteem, they expect the responses to be consistent across those items. For example, on the Rosenberg Self-Esteem Scale, someone who agrees with the statement "I feel I am a person of worth" should also agree with "I have a number of good qualities". If responses to these items are not correlated, it suggests they might not be measuring the same underlying trait.

This principle applies not only to self-report questionnaires but also to behavioural and physiological measures. For instance, if someone repeatedly places high bets in a simulated gambling game, their behaviour demonstrates consistency in risk-taking. If their betting patterns are unpredictable, it undermines the claim that the game measures their level of risk-seeking behaviour.

Assessing Internal Consistency

Just like test-retest reliability, internal consistency is evaluated by collecting and analysing data. One common method is the split-half correlation:

1. The items on the measure are split into two groups, such as even-numbered and odd-numbered items.
2. Scores are calculated separately for each group.
3. The correlation between these two sets of scores is then assessed.

For example, in Figure 8.2.2, we can see that if university students' scores on the even-numbered items of the Rosenberg Self-Esteem Scale are strongly correlated with their scores on the odd-numbered items (e.g., +0.88), it indicates good internal consistency. Generally, a split-half correlation of +0.80 or higher suggests strong internal consistency.

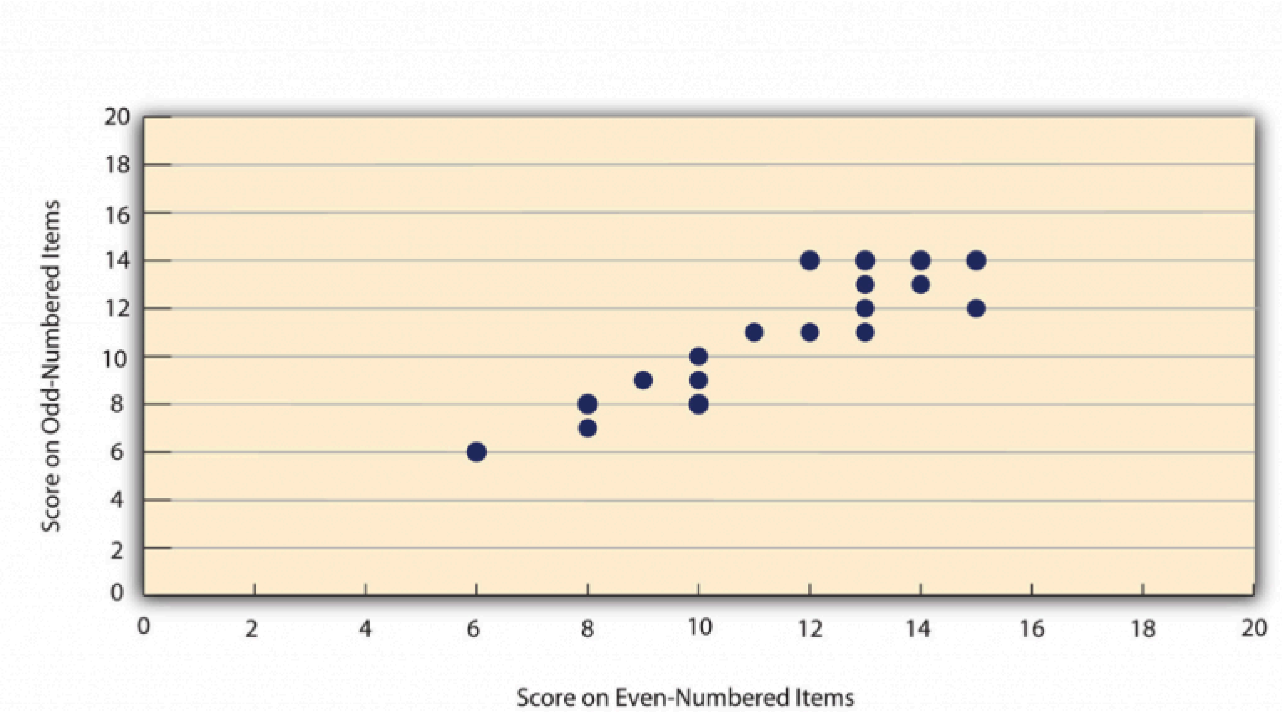


Figure 8.2.2. Split-half correlation between several college students' scores on the even-numbered items and their scores on the odd-numbered items of the Rosenberg Self-Esteem Scale by R. S. Jhangiani et al. in [Research Methods in Psychology 4e] is used under a [CC BY-NC-SA licence]

Cronbach's Alpha (α)

The most widely used statistic for measuring internal consistency is called Cronbach's alpha (α). Conceptually, Cronbach's α represents the average of all possible split-half correlations for a set of items. For example, if a survey has 10 items, there are 252 different ways to divide them into two groups of five items. Cronbach's α essentially summarises the average correlation across all these splits.

While the actual calculation of α is more complex, understanding it as an average of split-half correlations is a useful way to interpret its meaning. In general, an α value of +0.80 or higher indicates good internal consistency.

Inter-Rater Reliability

When researchers use behavioural measures to study human behaviour, they often rely on observers or raters to make judgements. Interrater reliability refers to how consistently different observers evaluate the same behaviour or event. In simple terms, it measures whether multiple observers agree in their assessments.

For example, imagine a study on university students' social skills. Researchers might record students interacting with someone they have just met and then ask two or more observers to watch the videos and rate each student's social skills. If social skills are a measurable trait, attentive observers should provide similar ratings for the same student. High agreement among observers indicates strong interrater reliability.

Another classic example is Bandura's Bobo doll experiment, where observers counted acts of aggression performed by children interacting with a clown-shaped doll. If two observers recorded similar numbers of aggressive actions for each child, their ratings would show high interrater reliability.

Measuring Interrater Reliability

To quantify interrater reliability, researchers use statistical tools:

- **Cronbach's alpha (α):** Used when the ratings are quantitative (e.g., rating social skills on a scale from 1 to 10).
- **Cohen's kappa (κ):** Used when the ratings are categorical (e.g., classifying behaviour as either "aggressive" or "non-aggressive").

Both statistics provide a numerical value that indicates the level of agreement among observers, helping researchers ensure that their measurements are reliable and consistent across different raters.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

8.3. VALIDITY

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Validity refers to how well a measurement tool captures the variable it is intended to measure. While reliability, which refers to the consistency of a measure, is an essential foundation for validity, it does not guarantee it. A measure can be highly reliable but lack validity entirely. For example, if someone attempted to measure self-esteem by using a ruler to measure index finger length, the results might be consistent every time (reliable) but would not reflect self-esteem (invalid). Therefore, researchers must gather evidence to support the claim that their measurement tools accurately represent the intended construct. This evidence is typically categorised into several types, including face validity, content validity, criterion validity, convergent validity, and discriminant validity.

Face Validity

Face validity refers to how much a measurement method appears to measure what it claims to measure, just by looking at it. For example, a self-esteem questionnaire with questions like “Do you see yourself as a person of worth?” or “Do you think you have good qualities?” would seem to have good face validity because these questions directly relate to self-esteem. On the other hand, trying to measure self-esteem by measuring someone’s finger length would have poor face validity because there is no obvious connection between the two.

Face validity is often evaluated informally. Researchers (or sometimes participants) simply consider whether the measure looks like it is assessing the intended construct. Occasionally, it is assessed quantitatively; for instance, by asking a group of people to rate how well they think a measure captures the concept it is supposed to measure.

However, face validity is one of the weakest forms of evidence for a measure’s accuracy. One reason is that it relies on intuition, and people’s assumptions about what should measure a construct are often incorrect. Many well-established psychological tests actually lack face validity but still work effectively.

For example, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is a widely used tool for assessing personality traits and disorders. It includes statements like “I enjoy detective or mystery stories” and “The sight of blood doesn’t frighten me or make me sick” to assess traits like aggression suppression.

At first glance, these items do not seem connected to aggression. However, the test is not concerned with individual answers to these questions but rather with how a person's overall pattern of responses compares to known patterns of individuals who suppress aggression.

Content Validity

Content validity refers to how well a measurement method captures all the important aspects of the construct it is supposed to measure. In other words, does the measure fully represent the concept it claims to assess?

For example, if a researcher defines test anxiety as including both physical symptoms (like nervous feelings caused by the activation of the sympathetic nervous system) and negative thoughts (like worrying about failure), then a good measure of test anxiety should include questions about both of these aspects. If the measure only focuses on nervous feelings and ignores negative thoughts, it would lack content validity because it does not fully represent the construct.

Similarly, attitudes are often described as including thoughts, feelings, and actions toward something. For instance, someone with a positive attitude toward exercise might:

- think positively about exercise ("Exercise is good for my health.")
- feel good about exercising ("I enjoy working out.")
- actually engage in exercise regularly.

A measure of attitudes toward exercise would need to include items assessing all three of these components to have strong content validity.

Unlike other forms of validity, content validity is not typically assessed through statistical analysis. Instead, it relies on a detailed comparison of the measure with the conceptual definition of the construct. Researchers carefully review whether the measure includes all relevant dimensions and adequately reflects the intended concept.

Criterion Validity

Criterion validity refers to how well a measurement correlates with other variables, called criteria, that are logically related to the construct being measured. In simple terms, it asks: Does this measure behave the way we expect it to when compared to other relevant outcomes?

For example, if a researcher develops a new test anxiety questionnaire, we would expect the scores on this measure to correlate negatively with performance on an important school exam. In other words, students with higher test anxiety scores should, on average, perform worse on the exam. If such a negative correlation is found, it suggests that the test anxiety measure is valid. However, if students with high anxiety scores perform just as well as those with low scores, it would raise doubts about whether the measure actually assesses test anxiety.

A criterion can be any variable that logically connects to the construct being measured. For test anxiety, relevant criteria might include:

- exam performance (expected negative correlation)
- overall course grades (expected negative correlation)
- blood pressure during an exam (expected positive correlation).

For another example, consider a measure of physical risk-taking. Validating such a measure might involve checking whether the scores are related to:

- participation in extreme sports like snowboarding or rock climbing
- the number of speeding tickets a person has received
- the number of broken bones they have experienced.

Types of Criterion Validity

- **Concurrent validity:** When the criterion is measured at the same time as the construct, it is called concurrent validity. For example, if test anxiety scores are correlated with blood pressure readings taken during an actual test, this would be evidence of concurrent validity.
- **Predictive validity:** When the criterion is measured in the future, it is called predictive validity. For example, if test anxiety scores predict lower final exam scores at the end of the semester, the measure demonstrates predictive validity.

Convergent Validity

A special case of criterion validity is called convergent validity. It examines whether scores on a new measure align with scores from existing, well-established measures of the same construct. If a new test anxiety scale correlates strongly with a widely accepted test anxiety questionnaire, this would demonstrate convergent validity.

For example, psychologists John Cacioppo and Richard Petty developed the Need for Cognition Scale, which measures how much people enjoy and value thinking (Cacioppo & Petty, 1982). To validate their scale:

- They found that higher scores correlated positively with academic achievement test scores.
- They found that higher scores correlated negatively with dogmatism (a measure of rigid, uncritical

thinking).

Over time, the Need for Cognition Scale has been shown to correlate with a variety of outcomes, such as the effectiveness of advertisements, political interest, and juror decision-making (Petty et al., 2009).

Discriminant Validity

Discriminant validity refers to the extent to which a measure does not strongly correlate with measures of variables that are theoretically different from it. In other words, it ensures that the measure captures the intended construct and not something else.

For example, self-esteem reflects a stable, long-term attitude toward oneself, while mood refers to temporary feelings that can change from moment to moment. A valid self-esteem questionnaire should show little correlation with a measure of mood. If the two are highly correlated, it could suggest that the self-esteem measure is unintentionally capturing mood rather than actual self-esteem.

When psychologists John Cacioppo and Richard Petty developed the Need for Cognition Scale (which measures how much people enjoy and value thinking), they also tested its discriminant validity. They found:

- only a weak correlation between need for cognition and cognitive style (e.g., whether someone tends to think analytically or holistically)
- no correlation between need for cognition and test anxiety
- no correlation between need for cognition and social desirability (the tendency to respond in a way that makes one appear socially acceptable).

These weak or nonexistent correlations provided strong evidence that the Need for Cognition Scale was measuring a distinct construct, separate from cognitive style, anxiety, or social desirability.

References

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

Petty, R., Briñol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition. In M. R. Leary & R. H. Hoyle, R. H. (Eds.). *Handbook of individual differences in social behavior.* The Guilford Press.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

8.4. PRACTICAL STRATEGIES FOR PSYCHOLOGICAL MEASUREMENT

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

When measuring a psychological construct for a research project, the process involves four key steps: defining the construct conceptually, defining it operationally, implementing the measure, and evaluating its effectiveness. Each step is crucial for ensuring that the measurement is both accurate and meaningful.

Conceptually Defining the Construct

The first step in measurement is to create a clear and precise conceptual definition of the construct. This definition serves as the foundation for every decision made about how the construct will be measured. For example, if a researcher wants to study "memory", a vague understanding of the term would lead to confusion about whether to measure memory for vocabulary words, visual images, or specific life events. Because psychologists now view memory as a collection of semi-independent systems, researchers need to pinpoint which type of memory they are investigating. If the focus is on episodic memory, which is memory for personal experiences, it would make sense to ask participants to recall events from last week. On the other hand, a task requiring participants to remember to complete an activity in the future would not align with this focus. Developing a strong conceptual definition often involves reviewing the existing research literature to understand how other experts have defined and measured the construct.

Operationally Defining the Construct

After defining the construct conceptually, the next step is to create an operational definition, which specifies exactly how the construct will be measured. Psychological constructs are often abstract and cannot be directly observed, so researchers must translate them into observable and measurable indicators. For example, stress could be measured through participants' scores on a stress scale, cortisol levels in saliva, or a tally of significant life events they have experienced recently.

In many cases, using an existing measure is the most efficient approach. Established measures have already undergone testing for reliability and validity, saving time and effort while also allowing results to be compared with prior studies. When choosing from multiple existing measures, researchers might select the one that is most commonly used, has the strongest evidence of validity, or best captures the specific aspect of the construct they are investigating.

For example, the Ten-Item Personality Inventory (TIPI) measures the Big Five personality traits with just ten questions. Although it is less reliable than longer inventories, it might be chosen when time is limited. Existing measures are often detailed in academic publications, or they may be catalogued in resources like the Directory of Unpublished Experimental Measures or PsycTESTS. Some widely used clinical tools, such as the Beck Depression Inventory or the MMPI, are proprietary and must be purchased from publishers.

If no suitable measure exists, researchers may choose to create their own. Developing a new measure often involves modifying existing tools, creating versions adapted for different formats (e.g., paper-based or digital), or repurposing tasks designed for other research contexts. For example, the Stroop Task, originally designed to measure cognitive control, has been adapted to study social anxiety by introducing socially charged words.

When designing a new measure, simplicity is key. Instructions should be clear and easy to understand, with opportunities for participants to ask questions before starting. Practice tasks can help participants become familiar with the procedure, and measures should be concise enough to prevent fatigue or loss of focus. However, brevity must be balanced with reliability. Single-item measures are often unreliable because responses can be influenced by misunderstandings, distractions, or random errors. Multiple-item measures are generally more reliable because they average out such inconsistencies.

Before fully implementing a new measure, it is wise to pilot-test it with a small group. Researchers can observe participants, track how long they take to complete the task, and gather feedback on clarity and difficulty. This trial phase allows researchers to address any problems before collecting large-scale data.

Implementing the Measure

The way a measure is administered plays a crucial role in its reliability and validity. Ideally, participants should be tested under similar conditions, preferably in a quiet, distraction-free environment. While group testing is often more efficient, it can introduce distractions or inconsistencies if not managed carefully. Researchers can minimise these risks by referring to previous studies that successfully used similar testing conditions.

Another challenge in implementation arises from participant reactivity, where individuals change their behaviour because they know they are being measured. Some participants may try to "please" the researcher by responding in socially desirable ways. For instance, someone with low self-esteem might claim to feel valuable simply because they believe it is the expected answer. Researchers must also consider demand characteristics, which are subtle cues in the study environment that hint at what behaviour is expected. For example, measuring attitudes about exercise immediately after showing participants an article about heart disease might unintentionally bias their responses.

To minimise these biases, researchers can guarantee anonymity, ensure participants cannot see each other's responses, and standardise the instructions given to all participants. Hypotheses and the true purpose of the study can also be concealed when appropriate. For example, a questionnaire titled "Financial Habits Survey" would be less likely to bias responses than one titled "Are You Financially Responsible?"

When possible, measures should be administered by someone "blind" to the study's hypothesis to prevent their expectations from unintentionally influencing participants. Consistency is key, and every participant should experience the study in the same way.

Evaluating the Measure

After data collection, researchers must evaluate the measure's reliability and validity to confirm it performed as expected. Even well-established measures need to be reassessed because new testing conditions or unique participant samples can affect results.

Reliability can be examined in several ways. Test-retest reliability can be assessed if participants completed the measure multiple times. For example, a professor might measure students' attitudes toward critical thinking at both the start and end of a semester. Even if attitudes remained stable, the correlation between the two sets of scores would provide insight into the measure's reliability over time. Similarly, internal consistency, often evaluated using statistics like Cronbach's alpha (α), ensures that multiple items within a measure are producing consistent results.

Validity is assessed by looking at how well the measure aligns with related variables. Criterion validity examines whether the measure correlates with expected outcomes. For instance, a mood scale should show distinct results between groups exposed to positive versus negative emotional stimuli, as demonstrated in MacDonald and Martineau's study on mood and self-esteem.

When reliability or validity falls short, researchers must consider possible explanations. Issues could stem from the measure itself, the way it was administered, or even the conceptual definition of the construct. If a mood scale shows no difference between participants instructed to think positive versus negative thoughts, it might mean the manipulation failed or the measure did not accurately capture mood changes. In such cases, adjustments must be made, whether by refining the measure, revising the conceptual definition, or improving the experimental design.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

8.5. SUMMARY

By Marc Chao

Summary

Measurement in psychology systematically assigns scores to individuals to represent traits or constructs, many of which, like intelligence or self-esteem, are intangible. Unlike physical attributes, these constructs require operational definitions to quantify them accurately through methods such as self-reports, behavioural observations, or physiological measures. For example, memory capacity might be measured using a backward digit span task, while stress could be assessed with questionnaires or biological markers like cortisol levels. These measurements range from nominal scales, which classify data into categories, to ratio scales, which include a true zero point and allow for precise comparisons. Reliability and validity are crucial to ensure that these measurements reflect their intended constructs accurately and consistently. Psychologists rigorously evaluate and refine their tools, much like questioning a faulty bathroom scale if its readings contradict observable changes, to ensure psychological research produces reliable and meaningful results.

Reliability in psychological measurement ensures consistency across various conditions, including time, items within a tool, and observers. Test-retest reliability assesses whether a measure produces stable results over time for stable traits, with high correlation coefficients (e.g., +0.80 or higher) indicating strong reliability. Internal consistency evaluates how well items within a multi-item measure align, often assessed using split-half correlations or Cronbach's alpha (α), where an α of +0.80 or higher suggests strong consistency. Interrater reliability measures the agreement between different observers, using tools like Cronbach's alpha for quantitative ratings or Cohen's kappa for categorical ones. These forms of reliability ensure that psychological tools are not only consistent but also credible, supporting the replicability of research findings and reinforcing trust in the results.

Validity, distinct from reliability, assesses whether a measurement tool accurately captures the construct it aims to measure. Face validity examines whether the measure appears to assess

the intended construct, though this is considered the weakest form of validity. Content validity ensures the measure comprehensively represents all aspects of the construct, closely aligning with its conceptual definition. Criterion validity evaluates how well the measure correlates with related outcomes and is divided into concurrent validity, where outcomes are measured simultaneously, and predictive validity, where outcomes are assessed in the future. Convergent validity ensures the measure aligns with established tools for the same construct, while discriminant validity confirms that it does not correlate with unrelated constructs. Together, these forms of validity provide a robust framework to evaluate the accuracy and appropriateness of psychological measurements.

The process of measuring a psychological construct involves four critical steps: defining the construct conceptually, defining it operationally, implementing the measure, and evaluating its effectiveness. A clear conceptual definition establishes what the construct represents, ensuring alignment between theory and measurement. The operational definition translates this abstraction into observable and measurable indicators, often using established tools like the Beck Depression Inventory or creating new measures suited to the study's needs. Implementation requires consistency in testing conditions to minimise biases such as participant reactivity or demand characteristics, ensuring the measure is both reliable and valid. Finally, researchers evaluate the measure's performance by reassessing its reliability (e.g., test-retest reliability or internal consistency) and validity (e.g., criterion or content validity). This iterative process allows for refinements in the measure, conceptual definition, or study design, ensuring accurate, meaningful results that advance psychological research.

CHAPTER 9: RESEARCH METHODS IN PSYCHOLOGY

Psychological research aims to explore and understand human behaviour, thoughts, and emotions through systematic investigation. The methodologies used in this field are broadly categorised into experimental and non-experimental approaches, each offering distinct strengths and limitations. Researchers choose the most appropriate method based on the nature of their research questions, ethical considerations, and practical constraints.

Experimental research is designed to establish cause-and-effect relationships by manipulating an independent variable (IV) and measuring its impact on a dependent variable (DV), while controlling for extraneous factors. Several experimental designs exist, including between-subjects designs, where different groups of participants experience different levels of the IV; within-subjects designs, where the same participants experience all levels of the IV; and matched-groups designs, where participants are matched on specific characteristics to ensure comparable groups. Quasi-experimental designs, which lack either random assignment or strict control over extraneous variables, are also commonly used in situations where traditional experimental methods are impractical.

In contrast, non-experimental methods do not involve manipulating variables. Instead, they focus on measuring and observing variables as they naturally occur to identify relationships or describe phenomena. These methods are particularly valuable when experiments are impractical, unethical, or impossible. Examples include correlational research, which examines the statistical relationships between variables; survey research, which gathers self-reported data through questionnaires or interviews; and qualitative research, which provides an in-depth exploration of human experiences and behaviours using non-numerical data.

Both experimental and non-experimental methods are essential for advancing psychological knowledge. While experimental methods are unparalleled in establishing causality, non-experimental methods excel in identifying patterns, generating hypotheses, and studying variables within natural contexts. The sections ahead will explore these methods in greater detail, highlighting their design, implementation, and interpretation, as well as their respective strengths and limitations. Whether conducted in controlled laboratory settings or through open-ended exploration, these research approaches are fundamental to enhancing our understanding of the human mind and behaviour.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand the role of experiments in establishing causal relationships:** Explain how independent and dependent variables are manipulated and measured in psychological experiments to determine cause-and-effect relationships.
- **Differentiate between between-subjects and within-subjects designs:** Compare the strengths and limitations of between-subjects and within-subjects experimental designs.
- **Explain the importance of controlling extraneous variables:** Describe how random assignment, standardisation, and counterbalancing help reduce the impact of extraneous variables on experimental outcomes.
- **Identify the four types of validity in research:** Define internal, external, construct, and statistical validity, and explain their significance in evaluating experimental research.
- **Differentiate between true experiments and quasi-experiments:** Explain the key differences between true and quasi-experiments and describe common quasi-experimental designs such as pretest-posttest and interrupted time series.
- **Understand the use of survey research in psychology:** Describe the key features of survey research, including self-report measures, sampling, and the design of effective survey questions.
- **Analyse factorial designs and interaction effects:** Explain the structure of factorial designs, identify main effects and interactions, and describe how factorial experiments allow researchers to examine complex relationships between multiple independent variables.
- **Compare qualitative, quantitative, and mixed-methods research:** Identify the strengths and limitations of qualitative, quantitative, and mixed-methods approaches in psychological research.

9.1. EXPERIMENT BASICS

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

In the late 1960s, social psychologists John Darley and Bibb Latané introduced a surprising idea: the more witnesses there are to an accident or crime, the less likely any one of them is to help the victim. This phenomenon, known as the "bystander effect", occurs because each witness feels less personal responsibility to take action, a process called "diffusion of responsibility".

Darley and Latané referenced real-world cases to support their theory, including the tragic murder of Catherine "Kitty" Genovese in New York. Despite multiple witnesses, no one intervened to stop the attack. However, the researchers recognised that single events like this could not conclusively prove their hypothesis. It was impossible to know whether fewer witnesses would have led to different outcomes.

To test their theory, Darley and Latané conducted a controlled experiment in a laboratory setting. Participants were isolated in separate rooms and told they would discuss university life over an intercom system. During the conversation, one of the supposed students (actually a pre-recorded voice) began to simulate an epileptic seizure, pleading for help with increasingly desperate and choking sounds.

Participants were placed in one of three conditions:

1. They believed they were the only witness to the emergency.
2. They thought one other person was also listening.
3. They assumed four other people were part of the discussion.

The results were striking: the more witnesses participants thought were present, the less likely they were to help. In the one-witness condition, 85% of participants sought help. This dropped to 62% with two witnesses and 31% with five.

This experiment is a classic example of how causal relationships between variables are tested in psychology. By carefully controlling conditions and isolating the variable of "number of witnesses", Darley and Latané demonstrated a clear cause-and-effect relationship. This study remains one of the most influential pieces of research in social psychology, illustrating the power of experiments to uncover how human behaviour is influenced by social situations.

What Is an Experiment?

An experiment is a research method designed to determine if there is a causal relationship between two variables. In other words, whether one variable directly causes changes in another. These two variables are referred to as the independent variable (the one that is manipulated) and the dependent variable (the one that is measured).

Experiments have two key characteristics.

First, researchers manipulate the independent variable by systematically changing its levels, which are known as conditions. For example, in Darley and Latané's study on helping behaviour, the independent variable was the number of witnesses participants believed were present. The researchers created three conditions by telling participants there were either one, two, or five other students in the discussion. It is important to note that while there are three conditions, there is still only one independent variable, namely the number of witnesses, with three levels (one, two, or five). Beginners sometimes mistakenly think that each condition represents a separate independent variable, but this is not the case.

Second, researchers control extraneous variables, meaning factors other than the independent or dependent variable that might influence the results. In their study, Darley and Latané ensured consistency by testing all participants in the same room and exposing them to the same emergency scenario. They also randomly assigned participants to each condition to ensure the groups were similar at the start of the experiment.

While the words "manipulation" and "control" might seem similar, they have distinct meanings in research. Researchers manipulate the independent variable by changing its levels, while they control extraneous variables by keeping them constant across conditions. Together, these two features allow experiments to provide clear, reliable evidence of cause-and-effect relationships.

Manipulating the Independent Variable

Manipulating an independent variable means systematically changing its levels so that participants are exposed to different conditions. This can happen either by assigning different groups of participants to different levels of the variable or by exposing the same group to different levels at different times.

For example, if a researcher wants to study whether expressive writing affects health, they might ask one group of participants to write about traumatic experiences and another group to write about neutral experiences. These two scenarios represent different levels of the independent variable, often referred to as conditions. Researchers typically name these conditions for clarity, such as the "traumatic condition" and the "neutral condition".

A key point is that manipulation requires active intervention by the researcher. Simply comparing groups that already differ on the independent variable is not manipulation and, therefore, is not an experiment.

For instance, if a researcher compares people who already keep journals with those who do not, they are not manipulating anything. Pre-existing groups like these might differ in other meaningful ways, such as personality traits (e.g., being more conscientious or introverted) or stress levels. As a result, any observed health differences could stem from these factors rather than from journaling itself. Active manipulation is essential for ruling out alternative explanations and establishing a clear cause-and-effect relationship.

However, there are situations where manipulating an independent variable is not possible due to practical or ethical reasons. For example, a researcher cannot randomly assign people to have or not have early childhood illnesses to study their effects on hypochondriasis. While experiments are not possible in such cases, researchers can still study these relationships using non-experimental methods, which will be discussed later.

Two-Level vs. Multi-Level Designs

Independent variables can be manipulated to create either two conditions or multiple conditions:

- **Single-Factor Two-Level Design:** The independent variable has two conditions (e.g., a group with one witness vs. a group with five witnesses).
- **Single-Factor Multi-Level Design:** The independent variable has more than two conditions (e.g., one witness, two witnesses, and five witnesses).

In Darley and Latané's bystander study, they used a single-factor multi-level design with three conditions: one witness, two witnesses, and five witnesses. This allowed them to observe patterns across multiple levels of the independent variable, offering richer insights than a simple two-level comparison would have provided.

Control of Extraneous Variables

Extraneous variables are any factors in a study that are not the independent or dependent variables but could still influence the results. In an experiment testing whether expressive writing affects health, examples of extraneous variables include:

- **Participant variables:** writing ability, diet, gender
- **Situational variables:** time of day participants write, whether they use a computer or write by hand, and even the weather

These variables are a problem because they can affect the dependent variable (e.g., participants' health) in ways unrelated to the independent variable (expressive writing). For instance, someone's health might improve because of their healthy diet rather than their participation in expressive writing exercises.

If extraneous variables are not controlled, they can make it hard to tell if changes in the dependent variable were caused by the independent variable or by something else.

To address this, researchers must control extraneous variables by holding them constant across all participants and conditions. For example:

- All participants could be asked to write at the same time of day.
- Everyone could use the same method of writing (e.g., all on a computer).

By keeping these variables consistent, researchers can reduce their influence and ensure that any observed effects on the dependent variable are more likely to be caused by the independent variable.

Extraneous Variables as "Noise"

Extraneous variables can make it harder to see the effect of an independent variable in two key ways. One major way is by adding "noise" or extra variability to the data, which can obscure meaningful results.

Imagine an experiment testing how mood (happy vs. sad) affects people's ability to recall happy childhood events. Participants are shown a happy or sad video clip to set their mood, then asked to recall as many happy memories as possible.

In a perfect world with no extraneous variables, the data would be clear. Every participant in the happy mood condition might recall exactly four happy memories, and every participant in the sad mood condition might recall exactly three memories. The difference is obvious.

In reality, extraneous variables create variability in the data. For example:

- Some participants in a happy mood might recall fewer happy memories because they have fewer to draw on or are less motivated.
- Some participants in a sad mood might recall more memories because they naturally have better recall strategies.

Even if the average difference between the groups stays the same, the increased variability (or noise) makes the effect of mood harder to detect.

How to Control Extraneous Variables

One way to reduce noise is to hold extraneous variables constant:

- **Situational variables:** Test all participants in the same room, give identical instructions, and treat everyone exactly the same.

- **Participant variables:** Some studies control for specific participant traits. For example, language studies often limit participants to right-handed people because brain areas related to language are usually more consistent in right-handed individuals.

Balancing Control and External Validity

In theory, researchers could control extraneous variables by selecting only a very specific group of participants, for example, 20-year-old, right-handed, female psychology majors. This would reduce variability but also limit external validity, which is the ability to apply the results to a broader population.

For instance:

- Results from young, female participants might not generalise to older male participants.

In most cases, researchers aim for a balance: they reduce variability enough to detect meaningful effects while keeping the sample diverse enough to ensure the findings can be applied to a wider group of people.

Extraneous Variables as Confounding Variables

Extraneous variables can interfere with an experiment by becoming confounding variables. A confounding variable is an extraneous variable that systematically varies along with the independent variable, making it difficult to determine which one is actually causing changes in the dependent variable.

For example, in most experiments, participants' IQ (intelligence quotient) is considered an extraneous variable because people naturally have different IQ levels. This variability is usually acceptable if IQ is evenly distributed across all experimental groups, meaning participants with lower and higher IQs are equally present in each condition. Figure 9.1.1 illustrates a hypothetical example: if participants in one group, such as a happy mood condition, have much higher IQs on average than those in another group, such as a sad mood condition, IQ becomes a confounding variable.

The word "confound" means to confuse, and that is precisely what confounding variables do. If participants in the happy mood condition perform better on a memory task, it becomes unclear whether their improved performance is due to their mood or their higher IQs. When two variables, such as mood and IQ, vary together, it becomes impossible to know which one is responsible for the observed outcome.

One way to avoid confounding variables is to hold them constant. For example, researchers could control IQ by only including participants with an IQ of exactly 100. While this approach prevents IQ from becoming a confounding variable, it significantly reduces diversity in the sample and limits how well the results can apply to a broader population.

A more practical and widely used approach is random assignment, where participants are randomly placed into experimental groups. Random assignment helps ensure that extraneous variables, such as IQ, are evenly distributed across conditions, reducing the likelihood that they will confound the results.

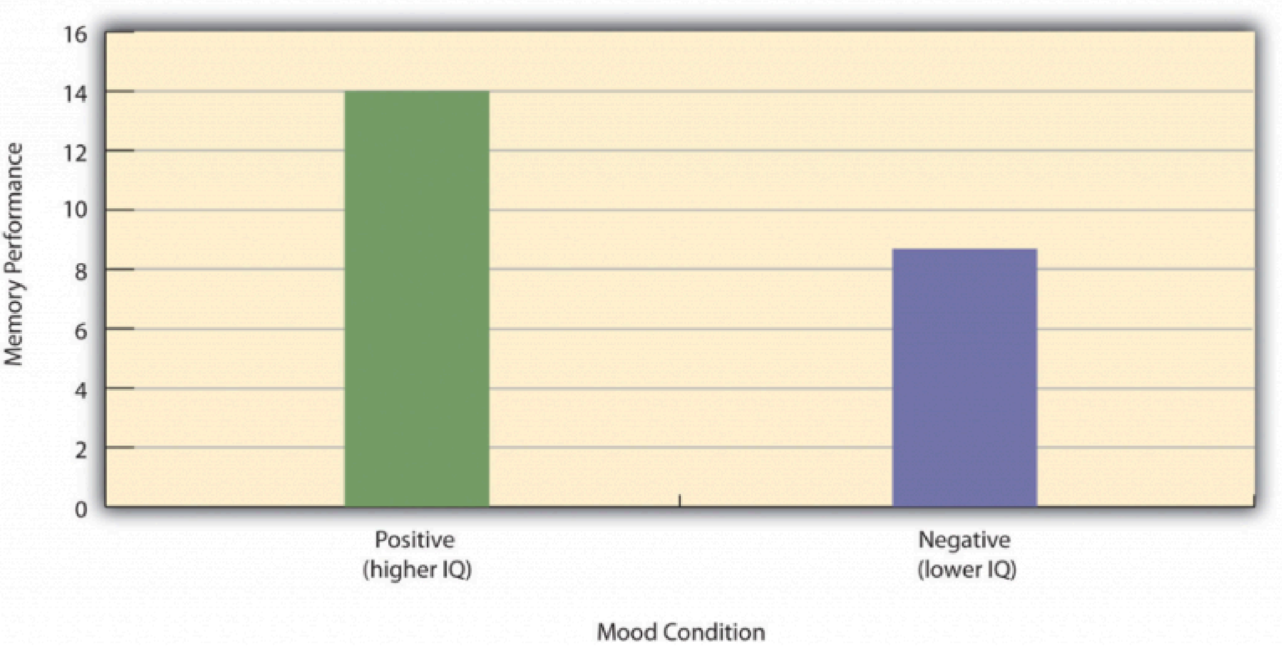


Figure 9.1.1. Hypothetical results from a study on the effect of mood on memory. Because IQ also differs across conditions, it is a confounding variable by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Treatment and Control Conditions

In psychological research, a treatment refers to any intervention designed to improve behaviour, mental health, or well-being. This can include therapies for mental health disorders, medical treatments, educational strategies, or programs aimed at reducing prejudice or promoting conservation. To test if a treatment works, participants are randomly assigned to either a treatment condition (where they receive the treatment) or a control condition (where they do not). If participants in the treatment condition show better outcomes, such as reduced depression, faster learning, or improved behaviour, researchers can conclude that the treatment is effective. Studies like these, especially in medical or therapeutic contexts, are often called randomised clinical trials.

Control conditions help researchers ensure that improvements in the treatment group are due to the treatment itself and not other factors. In a no-treatment control condition, participants receive no treatment at all. However, this approach has a limitation: placebo effects. People often experience improvement simply because they expect to feel better. This expectation can reduce stress, anxiety, and even improve immune responses, creating real changes despite the absence of an actual treatment.

Placebo effects are fascinating but problematic for research. Imagine a study where participants in a treatment group improve more than those in a no-treatment control group. Figure 9.1.2 illustrates hypothetical results where participants in a treatment condition show greater improvement, on average, compared to those in a no-treatment control condition. However, if these two conditions (represented by the two leftmost bars in Figure 9.1.2) were the only comparison points, it would be impossible to confidently conclude that the observed improvement was due to the treatment itself. It remains possible that participants in the treatment group improved primarily because they expected to do so, while those in the no-treatment control group lacked this expectation.

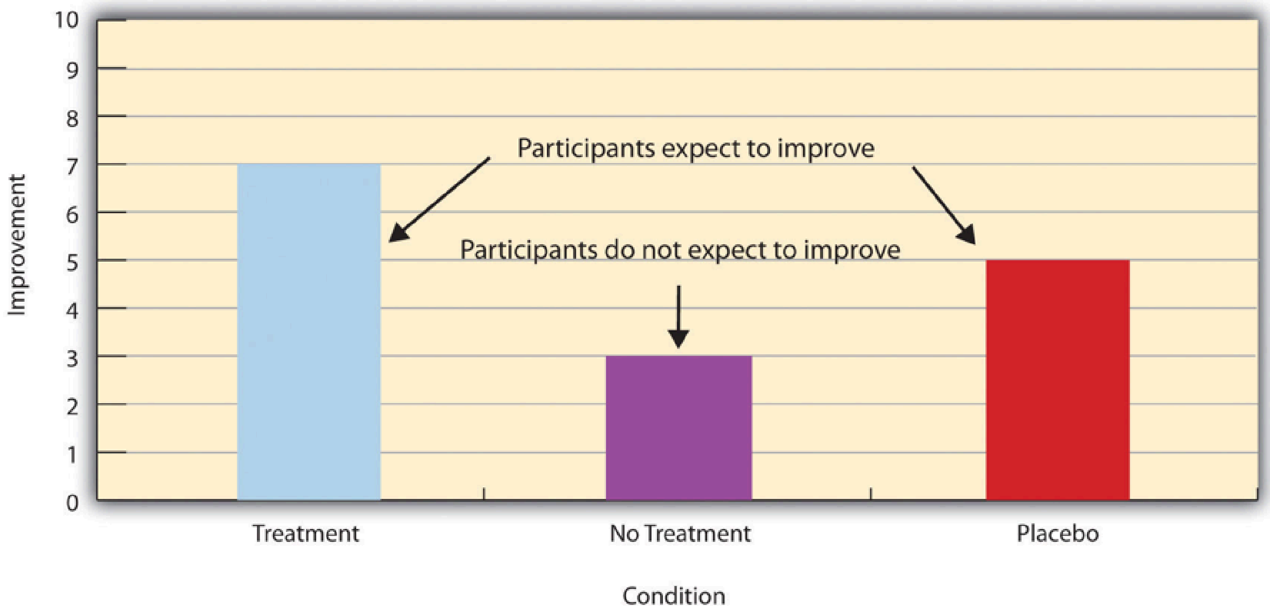


Figure 9.1.2. Hypothetical results from a study including treatment, no-treatment, and placebo conditions by R. S. Jhangiani et al. in [Research Methods in Psychology 4e] is used under a [CC BY-NC-SA licence]

One way to address this is to include a placebo control condition. In this setup, participants receive something that looks and feels like the real treatment but lacks its active component. For example, if the treatment involves taking a pill, participants in the placebo group would take an identical pill without the active ingredient. In psychotherapy research, the placebo might involve unstructured talk therapy, where participants meet a therapist but without any structured therapeutic techniques. If both groups expect improvement, but the treatment group still shows greater improvement, researchers can confidently attribute the effect to the treatment itself.

Ethical guidelines require participants to be informed about whether they might receive a treatment or a placebo, although they will not know which one until the study ends. Often, participants in the placebo group are offered the real treatment once the study is over.

Another option is a wait-list control condition. Participants in this group know they will eventually receive the treatment but must wait until others have completed it first. This setup allows researchers to compare people currently receiving the treatment with those still waiting, but who also expect improvement in the future.

Finally, researchers might skip a control condition entirely and compare a new treatment with the best existing alternative treatment. For example, a new therapy for phobias could be compared with standard exposure therapy. Since both groups receive active treatment, their expectations for improvement are similar. This approach helps answer a more meaningful question: "Is the new treatment better than what's already available?" instead of just "Does it work at all?"

The Power of the Placebo Effect

Many people expect placebos to help with psychological conditions like depression, anxiety, or insomnia. However, placebos have also been shown to improve conditions generally considered purely physical, such as asthma, ulcers, and even warts (Shapiro & Shapiro, 1999). Remarkably, there is evidence that placebo surgery, which is also known as "sham surgery", can sometimes be as effective as actual surgical procedures.

In a notable study, medical researcher J. Bruce Moseley and his colleagues investigated the effectiveness of two arthroscopic surgery procedures for osteoarthritis of the knee (Moseley et al., 2002). In the control group, participants underwent the full pre-surgical process: they were prepped, given a tranquilliser, and even had three small incisions made on their knees. However, they did not actually receive the arthroscopic procedure.

It is important to note that the study's use of deception would have undergone a thorough review by an Institutional Review Board (IRB). The IRB determined that the potential benefits of the study outweighed the risks and that no other method could have answered the research question about placebo procedures effectively.

The results were striking. All participants, including those in the placebo (sham surgery) group, showed improvements in both knee pain and function. In fact, the placebo group improved just as much as those who received the actual surgical procedure. As the researchers concluded, "This study provides strong evidence that arthroscopic lavage with or without débridement [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" (p. 85).

References

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

Shapiro, A. K., & Shapiro, E. (1997). *The powerful placebo: From ancient priest to modern physician*. Johns Hopkins University Press.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

9.2. EXPERIMENTAL DESIGN

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

This section explores different ways to design an experiment. The main difference lies in how participants interact with the independent variable. In one approach, each participant experiences only one level of the independent variable. This is known as a between-subjects experiment. In the other approach, each participant experiences every level of the independent variable. This is called a within-subjects experiment.

Between-Subjects Experiments

In a between-subjects experiment, each participant is exposed to only one level of the independent variable. For example, in a study with 100 university students, half might be asked to write about a traumatic event while the other half writes about a neutral event. Similarly, in a study with 60 people who have severe agoraphobia, 20 participants might be assigned to each of three different treatments for the disorder.

In this type of experiment, it is crucial that the groups are, on average, as similar as possible. Participants in each condition should have comparable characteristics, such as gender balance, average IQ, motivation levels, and general health status. This similarity ensures that extraneous participant variables, which are factors other than the independent variable, do not become confounding variables that could distort the results. By carefully balancing these variables across groups, researchers can be confident that any observed differences in outcomes are due to the independent variable, not unintended factors.

Random Assignment

Random assignment is a method researchers use to evenly distribute extraneous variables across different experimental conditions. It involves assigning participants to conditions using a random process, ensuring each participant has an equal chance of being placed in any group.

It is important not to confuse random assignment with random sampling. Random sampling is about selecting participants from a population, while random assignment focuses on distributing participants into experimental groups. In psychology, random sampling is rarely used, but random assignment is a standard and crucial practice.

For random assignment to be effective, two criteria must be met:

1. **Equal Chance:** Every participant must have the same probability of being assigned to each condition (e.g., a 50% chance for two conditions).
2. **Independence:** Each participant's assignment must be made independently of others.

A simple example of random assignment is flipping a coin: heads could mean the participant goes to Condition A, while tails assigns them to Condition B. For three conditions, researchers might use a random number generator to assign participants based on numbers (e.g., 1 for Condition A, 2 for Condition B, 3 for Condition C).

In practice, researchers often create an assignment sequence in advance, especially when using software. The sequence ensures that participants are assigned fairly as they arrive.

Addressing Unequal Group Sizes

A challenge with pure random assignment (e.g., coin flipping) is that group sizes might become unequal. While unequal sample sizes are not usually a big issue, equal group sizes are more efficient for statistical analysis.

To address this, researchers often use a method called block randomisation. In this approach:

- Each condition appears once within a block before any condition is repeated.
- The order of conditions within each block is randomised.
- This sequence is prepared before participants arrive, and each new participant is assigned to the next available slot in the sequence.

For example, if there are three conditions (A, B, and C) and nine participants, the random assignment might look like this (Table 9.2.1):

Table 9.2.1. Block randomisation 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**

Online tools, such as Research Randomiser, can help generate these sequences automatically.

Limitations of Random Assignment

While random assignment is highly effective, it is not without its limitations. There is always a chance that, by pure coincidence, the groups might differ in meaningful ways. For example, one group could unintentionally have slightly older participants, or participants who are more motivated than those in another group.

However, this concern is generally minimal for several reasons. First, random assignment tends to work better with larger sample sizes, as larger groups reduce the impact of chance imbalances. Second, statistical tests used to analyse experimental data are specifically designed to account for the imperfections of random assignment. Finally, if an unnoticed confounding variable does influence the results, replication of the experiment can often reveal and address such issues.

Matched Groups

In a matched-groups design, participants are carefully matched across conditions based on their scores on the dependent variable or other relevant extraneous variables before the independent variable is manipulated. This approach ensures that these variables will not become confounding factors across experimental conditions.

For example, imagine we want to study whether expressive writing impacts people's health. First, we would measure health-related variables for all potential participants. Using these measurements, we would rank participants from the healthiest to the least healthy.

Next, we would pair participants based on their health rankings. In each pair, one participant would be randomly assigned to the traumatic writing condition, while the other would be assigned to the neutral writing condition. This process would continue until every participant is assigned to a condition, ensuring that both groups are balanced in terms of health from the start.

If we observe a difference in health outcomes between the two groups at the end of the study, we can confidently attribute this difference to the writing intervention rather than pre-existing differences in health. This design reduces variability between groups and strengthens the internal validity of the experiment.

Within-Subjects Experiments

In a within-subjects experiment, each participant experiences all conditions of the study. For example, in an experiment examining how a defendant's physical attractiveness influences judgements of guilt, a between-subjects design would involve one group evaluating an attractive defendant and another group evaluating an unattractive defendant. In contrast, a within-subjects design would have the same participants evaluate both an attractive and an unattractive defendant.

The main advantage of a within-subjects design is its ability to control extraneous participant variables effectively. Since each participant serves as their own control, factors like IQ, socioeconomic status, or family background remain consistent across conditions. This reduces variability caused by individual differences and makes it easier to detect the effect of the independent variable. Additionally, within-subjects designs allow researchers to use statistical techniques that account for these consistent participant variables, further minimising "noise" in the data.

However, not all experiments are suitable for a within-subjects design, and in some cases, it may not be the best choice. Certain types of studies or research questions might require a different approach, which we will explore further later in the chapter.

Carryover Effects and Counterbalancing

In a within-subjects experiment, participants experience all conditions of the independent variable. While this design has advantages, it also introduces potential issues known as order effects. Order effects happen when the order in which participants experience conditions influences their responses. One common type is a carryover effect, where being tested in one condition affects performance in later conditions.

Carryover effects can take different forms. A practice effect occurs when participants perform better in later conditions because they have had time to practice a task. In contrast, a fatigue effect happens when participants perform worse in later conditions due to tiredness or boredom. Another type, called a context effect (or contrast effect), happens when participants' perceptions are influenced by the order of conditions.

For example, if participants first judge an attractive defendant and then judge an average-looking one, their judgements might be harsher simply because of the comparison.

Order effects can also make participants more likely to guess the hypothesis of the study. If participants are asked to judge both an attractive and an unattractive defendant, they might realise the study is about how attractiveness influences judgements. This knowledge could lead them to unconsciously alter their responses to match (or oppose) the perceived expectations of the researcher.

While carryover effects are sometimes interesting research topics on their own, they pose a problem when they are not the focus of the study. For example, if participants always judge the attractive defendant first and the unattractive defendant second, any differences observed might result from the order of conditions rather than from the attractiveness variable itself. In this case, the order of conditions becomes a confounding variable.

The most effective solution to this issue is counterbalancing, where participants experience conditions in different orders. The best approach is complete counterbalancing, where every possible order of conditions is used equally across participants. For example, in a study with two conditions (A and B), half the participants would experience condition A first, while the other half would experience condition B first. With three conditions (A, B, and C), participants would be randomly assigned to one of six possible orders: ABC, ACB, BAC, BCA, CAB, or CBA. However, as the number of conditions increases, the number of possible orders grows rapidly. For example, four conditions require 24 orders, and five conditions require 120 orders.

When complete counterbalancing is not practical, researchers often use a Latin square design. A Latin square ensures that each condition appears in every position (first, second, third, etc.) an equal number of times and that every condition follows and precedes each other condition exactly once. For example, in a study with four conditions (A, B, C, D), the Latin square might look like this (Table 9.2.2):

Table 9.2.2. An example of a Latin square

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

In this setup, each condition appears in every order position once, and the total number of orders equals the number of conditions (4 instead of 24). This approach drastically reduces the complexity of counterbalancing while still minimising order effects.

For experiments with a very large number of conditions, random counterbalancing can be used. In this method, the order of conditions is randomly assigned for each participant. While this approach is less effective than complete counterbalancing or Latin square designs, it can still help reduce order effects when they are expected to be minor.

Counterbalancing achieves two key objectives. First, it prevents order effects from becoming confounding variables by ensuring conditions are presented in different orders across participants. This way, any observed differences in the dependent variable cannot be solely attributed to the order of conditions. Second, counterbalancing allows researchers to detect carryover effects by analysing whether the order of conditions had any significant impact on the results.

Simultaneous Within-Subjects Designs

In traditional within-subjects designs, participants experience one condition at a time. However, there is an alternative approach where participants respond to multiple conditions simultaneously. This method is often used when participants make repeated responses in each condition.

For example, imagine a study where participants judge the guilt of 10 attractive defendants and 10 unattractive defendants. Instead of having participants rate all the attractive defendants first and then all the unattractive ones, the researcher could mix the two types together in a random order. Participants would then make judgements for all 20 defendants, and the researcher could calculate the average guilt rating for each type.

Another example might involve studying memory in people with social anxiety disorder. Suppose a researcher wants to know whether these individuals remember negative adjectives (e.g., “stupid”, “incompetent”) better than positive ones (e.g., “happy”, “productive”). Instead of presenting two separate lists of positive and negative adjectives, the researcher could create one list containing both types of words. Participants would study the mixed list and then try to recall as many words as possible. The researcher would then count how many positive and negative words were remembered.

This simultaneous approach allows participants to process multiple conditions within a single session, which can help reduce order effects and make the study more efficient. It also provides a clear comparison between conditions while maintaining the advantages of a within-subjects design.

Choosing Between Between-Subjects and Within-Subjects Designs

Most experiments can be designed using either a between-subjects or a within-subjects approach. Researchers must carefully consider the strengths and weaknesses of each method to determine which is best suited for their specific study.

Between-subjects designs are often simpler to set up and require less time per participant. They naturally avoid carryover effects without the need for complex counterbalancing. This design is particularly useful when testing time is limited or when exposure to one condition might permanently alter a participant's response in another condition.

On the other hand, within-subjects designs offer better control over extraneous participant variables, reducing noise in the data and making it easier to detect the effect of the independent variable on the dependent variable. They also typically require fewer participants to achieve the same statistical power as a between-subjects design.

A good general rule is: if you can conduct a within-subjects experiment in the available time per participant, and if carryover effects can be managed with proper counterbalancing, then a within-subjects design is usually the better choice. However, if a within-subjects design is impractical, either because of time constraints or because one condition might permanently affect responses in another, a between-subjects design is more appropriate.

For example, if you are studying participants in a busy setting like a doctor's waiting room or a grocery store line, you may not have the time to test each person under multiple conditions. A between-subjects design would be more efficient in this case. Similarly, if you are testing an intervention designed to reduce prejudice, a within-subjects design would require participants to be exposed to the treatment and then the control condition. If the treatment is effective, participants' prejudice levels would already be reduced, making them unsuitable for the control condition. In such cases, a between-subjects design is the only feasible option.

Finally, remember that choosing one design does not exclude the other from future studies. Researchers often use both designs, sometimes even within the same research program, to explore a question from multiple angles. This mixed-methods approach is common in professional research and can provide a more comprehensive understanding of the phenomenon being studied.

Chapter Attribution

Content adapted, with editorial changes, from:

9.3. EXPERIMENTATION AND VALIDITY

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

The Four Big Validities in Psychology Research

When evaluating a psychology experiment, one key question to ask is: "Is this study valid?" However, determining validity is not as simple as it might seem because there are different types of validity, each addressing a specific aspect of a study's accuracy and soundness.

Researchers generally focus on four main types of validity to determine whether an experiment is well-designed (Judd & Kenny, 1981; Morling, 2014):

1. **Internal Validity:** Does the study establish a clear cause-and-effect relationship between the independent and dependent variables?
2. **External Validity:** Can the study's results be generalised to other people, settings, or situations?
3. **Construct Validity:** Does the study accurately measure the concepts or variables it claims to measure?
4. **Statistical Validity:** Are the statistical analyses appropriate, and do they support the study's conclusions?

Each of these validities focuses on a different question about the research. In the following sections, we will take a closer look at each type to better understand how they contribute to a study's overall quality and reliability.

Internal Validity: Ensuring Cause-and-Effect Relationships

Just because two variables are statistically related does not mean one causes the other. You have probably heard the saying, "Correlation does not imply causation". For example, if studies show that people who exercise regularly tend to be happier, it does not automatically mean that exercise causes happiness. It is possible that happier people are more likely to exercise, or that another factor, such as better physical health, leads to both increased happiness and regular exercise.

The goal of an experiment is to demonstrate a causal relationship between two variables by showing that changes in the independent variable directly cause changes in the dependent variable. The logic is straightforward: if a researcher creates two or more similar conditions and only manipulates the independent variable while keeping everything else constant, then any differences observed in the dependent variable must have been caused by the independent variable.

Take Darley and Latané's experiment as an example. The only difference between their experimental conditions was the number of students participants believed were involved in the discussion. Because this was the only manipulated difference, it must have been the cause of differences in helping behaviour across the conditions.

A study is said to have high internal validity if its design supports the conclusion that the independent variable caused the observed changes in the dependent variable. Experiments typically have strong internal validity because they involve direct manipulation of the independent variable and control of extraneous variables, often through techniques like random assignment.

In contrast, non-experimental designs (e.g., correlational studies), where variables are observed and measured but not directly manipulated by the researcher, generally have lower internal validity. Without manipulation and control, it is harder to rule out alternative explanations for observed relationships between variables.

External Validity: How Well Do Results Apply Beyond the Experiment?

Experiments often face criticism for being conducted under artificial conditions due to the need to manipulate independent variables and control extraneous factors. For example, in many psychology experiments, participants are typically undergraduate students who complete paper-and-pencil questionnaires or computer tasks in a laboratory setting. Consider a study by Barbara Fredrickson and her colleagues (1998), where undergraduate students were asked to complete a maths test while wearing swimsuits. At first glance, this scenario seems unrealistic, as it is difficult to imagine a situation outside of a laboratory where someone would need to solve maths problems in a swimsuit.

This concern highlights the issue of external validity, which refers to how well the results of a study can be generalised to people and situations beyond those directly studied. A study has high external validity when its participants and conditions closely match real-world scenarios, a concept known as mundane realism. For example, if researchers wanted to study how cereal box colours (yellow vs. purple) influence shoppers' choices, they would achieve high external validity by observing real shoppers in a real grocery store. If shoppers bought more cereal in purple boxes, the findings would likely apply to other grocery stores and shoppers.

However, if the same study were conducted in a university lab, where undergraduate students simply rated colours on a computer screen, the study would have lower mundane realism. While the visual processing of colours might still reflect real-world decision-making (psychological realism), the results would not directly translate to real grocery store behaviour.

It is important to note that experiments are not inherently low in external validity. Many experiments are carefully designed to simulate real-world conditions. For example, Darley and Latané's experiment realistically simulated an emergency situation. Additionally, field experiments, which take place outside the lab, often achieve high external validity. In one study, Robert Cialdini and his colleagues tested how hotel guests responded to different towel reuse messages. They found that guests were far more likely to reuse their towels when told that most other guests did the same. Since the experiment was conducted in real hotel rooms with real guests, the findings are highly generalisable to other hotels.

Another reason experiments can still have strong external validity is that they often focus on universal psychological processes, which are mechanisms that operate consistently across different people and situations. Returning to Fredrickson's swimsuit study, the researchers found that women, but not men, performed worse on the maths test while wearing swimsuits. They concluded that this was due to self-objectification, where women are more likely to view themselves from an outsider's perspective, which can divert attention away from other tasks. While solving maths problems in swimsuits might be rare, the underlying psychological process of self-objectification is likely to occur in many different situations and contexts.

Construct Validity: How Well Does the Experiment Measure What it Claims to Measure?

Construct validity refers to how well an experiment's design captures the concept it intends to study. It focuses on whether the manipulations and measures accurately represent the research question.

In their famous study, Darley and Latané explored the question: "Does helping behaviour become diffused when more people are present?" They hypothesised that participants would be less likely to help in an emergency if they believed more people were available to assist. This process of translating a research question into an experimental design is called operationalisation, which refers to defining how abstract variables will be measured or manipulated.

Darley and Latané operationalised "diffusion of responsibility" by varying the number of other people participants believed were involved in the discussion. Their experiment created a clear emergency situation, provided participants with an opportunity to help, and systematically increased the number of perceived bystanders. This design had high construct validity because the manipulations closely aligned with the core research question.

However, what if the study had only included two conditions, one with a single student and another with two students in the discussion? While a decrease in helping behaviour might still be observed, it would not provide strong evidence for diffusion of responsibility. Instead, the effect might be interpreted as social inhibition, a concept from Bandura's research. In this case, construct validity would be lower because the study would not fully capture the intended phenomenon.

On the other hand, imagine if there had been five conditions instead of three. Researchers might have observed whether helping behaviour continued to decline as the number of bystanders increased or if it plateaued at a certain point. This would offer a more nuanced understanding of diffusion of responsibility. However, adding even more conditions beyond this point would not necessarily improve construct validity further, and it could make the study unnecessarily complex without adding new insights.

When designing your own experiment, think carefully about how well your operationalisation aligns with your research question. The goal is to ensure that your manipulations and measurements are clear, relevant, and directly address the concept you are trying to study. High construct validity strengthens the connection between your experimental design and the conclusions you can draw from your results.

Statistical Validity: Ensuring Accurate and Appropriate Data Analysis

Statistical validity refers to whether the correct statistical methods were used to analyse the data and if the conclusions drawn from those analyses are sound. In psychology, there are many statistical tests, such as *t*-tests, ANOVA, regression, and correlation, and choosing the right one depends on two main factors: the type of data collected (e.g., numerical or categorical) and the study design (e.g., between-subjects or within-subjects).

Each statistical test also comes with specific assumptions, like data being normally distributed or having equal variances between groups. If these assumptions are violated and the test is still applied, the statistical conclusions may become unreliable, which threatens statistical validity.

One common critique in research is that a study does not have enough participants. While this might seem like a concern about external validity (how well results generalise to a larger population), it is actually an issue of statistical validity. Small sample sizes make it harder to detect meaningful effects, even if they exist, because they reduce the statistical power of the study. However, it is worth noting that small sample sizes are not always a problem; certain types of research (e.g., single-case studies) can still provide valuable insights, as we will discuss later.

To ensure statistical validity, researchers must use the appropriate statistical tests for their data and design. Additionally, they should conduct a power analysis before starting their study. A power analysis helps determine the minimum number of participants needed to detect a specific effect size. In short, larger sample sizes increase the likelihood of detecting a real effect, but only if the right statistical tools are applied correctly.

Prioritising Validities in Research

The four key types of validity (internal, external, construct, and statistical) are essential tools for evaluating and designing experiments. However, achieving high validity across all four areas is often challenging, and researchers must prioritise based on their study's goals.

For example, in Cialdini's study on towel reuse in hotels, the external validity was notably high because the research was conducted in a real-world setting with typical hotel guests. However, the statistical validity was somewhat limited. This difference does not mean the study was flawed; instead, it highlights areas for improvement in future follow-up research (Goldstein et al., 2008).

As Morling (2014) explains, many psychology experiments tend to prioritise internal and construct validity to ensure that the independent variable truly causes changes in the dependent variable and that the variables are well-measured. However, this focus sometimes comes at the expense of external validity, meaning the findings may not always generalise well to real-world situations.

References

Cialdini, R. (2005). Don't throw in the towel: Use social influence research. *Observer, 2005*(April). https://www.psychologicalscience.org/observer/dont-throw-in-the-towel-use-social-influence-research

Fredrickson, B. L., Roberts, T.-A., Noll, S. M., Quinn, D. M., & Twenge, J. M. (1998). That 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. *The Journal of Consumer Research*, *35*(3), 472–482. https://doi.org/10.1086/586910

Judd, C.M. & Kenny, D.A. (1981). Estimating the effects of social interventions. Cambridge University Press.

Morling, B. (2014). Guide your students to become better research consumers. *Observer, 2014*(April). https://www.psychologicalscience.org/observer/teach-your-students-to-be-better-consumers

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

9.4. CORRELATIONAL RESEARCH

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Correlational research is a type of non-experimental study where researchers measure two variables (either binary or continuous) and analyse the statistical relationship between them, known as a correlation. In this type of research, there is little to no control over extraneous variables.

Researchers often choose correlational research when their goal is to describe and predict relationships between variables rather than establish causation. For example, if two variables are correlated, researchers can use regression analysis to predict one variable based on the other. This approach aligns with two primary goals of science: to describe phenomena and to make predictions about future outcomes.

In some cases, correlational research is preferred because manipulating an independent variable might be impossible, impractical, or unethical. For instance, studying the relationship between cannabis use frequency and memory performance cannot ethically involve manipulating how often people use cannabis. Instead, researchers measure both variables and analyse the correlation to understand their relationship without direct intervention.

Another significant advantage of correlational research is its role in establishing reliability and validity for measurement tools. For example, if a researcher develops a short extraversion questionnaire, they can compare it with a longer, well-established version. A strong correlation between the two sets of scores would suggest that the shorter test is valid and reliable.

Correlational studies also tend to have higher external validity compared to experiments. While experiments excel at internal validity through tight control over variables, they often create artificial conditions that might not exist in the real world. In contrast, correlational research is typically conducted in more natural settings, making its findings more generalisable to everyday situations.

Additionally, correlational research can complement experimental findings, providing converging evidence to strengthen a theory. When an experimental study demonstrates causation and a correlational study supports the same relationship in a natural context, researchers can have greater confidence in the theory's validity. For example, studies showing a correlation between watching violent television and aggressive behaviour have been reinforced by experiments that establish a causal connection (Bushman & Huesmann, 2011).

Does Correlational Research Always Involve Quantitative Variables?

A common misunderstanding among new researchers is the belief that correlational research must always involve two quantitative variables, such as scores from personality tests or the number of daily stressors and symptoms someone experiences. However, what truly defines correlational research is not the type of variables used, but rather the method of data collection, specifically, that both variables are measured and not manipulated. This principle holds true whether the variables are quantitative (e.g., test scores) or categorical (e.g., nationality or occupation).

For example, consider a study where a researcher administers the Rosenberg Self-Esteem Scale to 50 American college students and 50 Japanese college students. At first glance, this might seem like a between-subjects experiment, but it is actually correlational research because the researcher did not actively manipulate nationality; it was simply measured. Similarly, in the study by Cacioppo and Petty comparing college faculty and factory workers on their need for cognition, the research was correlational because participants' occupations were not manipulated, only observed.

To clarify this distinction, consider a hypothetical study investigating the relationship between making daily to-do lists and stress levels, such as the one shown in Figure 9.4.1. Whether this study is classified as correlational or experimental depends entirely on how it was conducted.

- If participants were randomly assigned to either make daily to-do lists or not make them, then this would be an experiment. In this case, researchers could make a causal claim, such as concluding that making daily to-do lists reduces stress.
- On the other hand, if participants were simply asked whether they usually make daily to-do lists, it would be a correlational study. In this scenario, researchers could only identify a statistical relationship between the two variables. They could not determine whether stress reduces people's ability to plan ahead (the directionality problem) or whether a third variable, such as conscientiousness, influences both stress levels and the likelihood of making to-do lists (the third-variable problem).

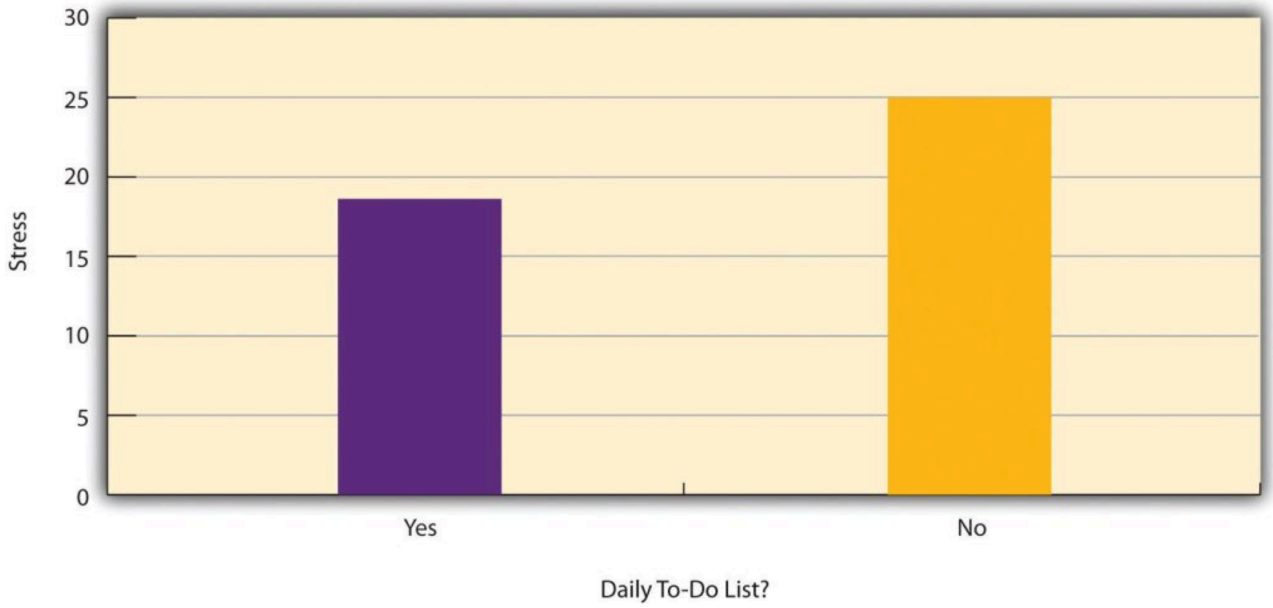


Figure 9.4.1. Results of a hypothetical study on whether people who make daily to-do lists experience less stress than people who do not make such lists by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Data Collection in Correlational Research

In correlational research, the key feature is that neither variable is manipulated by the researcher. What matters is that both variables are measured as they naturally occur, not how or where the data collection happens.

For example, a researcher might invite participants into a laboratory to complete a computerised backward digit span task (to measure memory) and a risky decision-making task (to measure risk tolerance). They could then analyse the relationship between the scores on these two tasks.

Alternatively, another researcher might visit a shopping mall and ask people about their attitudes toward the environment and their shopping habits. They would then assess the relationship between these two variables.

In both examples, even though the settings and data collection methods differ, the studies are still considered correlational because no variable was manipulated, they were only measured.

Correlations Between Quantitative Variables

Correlations between quantitative variables are often illustrated using scatterplots. In a scatterplot, each point represents one person's score on two variables. For example, imagine a study measuring the relationship between stress levels and physical symptoms. A point on the scatterplot might represent a person with a stress score of 10 and three physical symptoms (as shown in Figure 9.4.2). When observing all the points together, a pattern often emerges. If higher stress levels correspond with more physical symptoms, this indicates a positive relationship, as one variable increases, the other tends to increase as well. In contrast, a negative relationship occurs when higher scores on one variable are associated with lower scores on the other. For example, stress levels and immune system functioning typically show a negative relationship, where higher stress is linked to lower immune function.

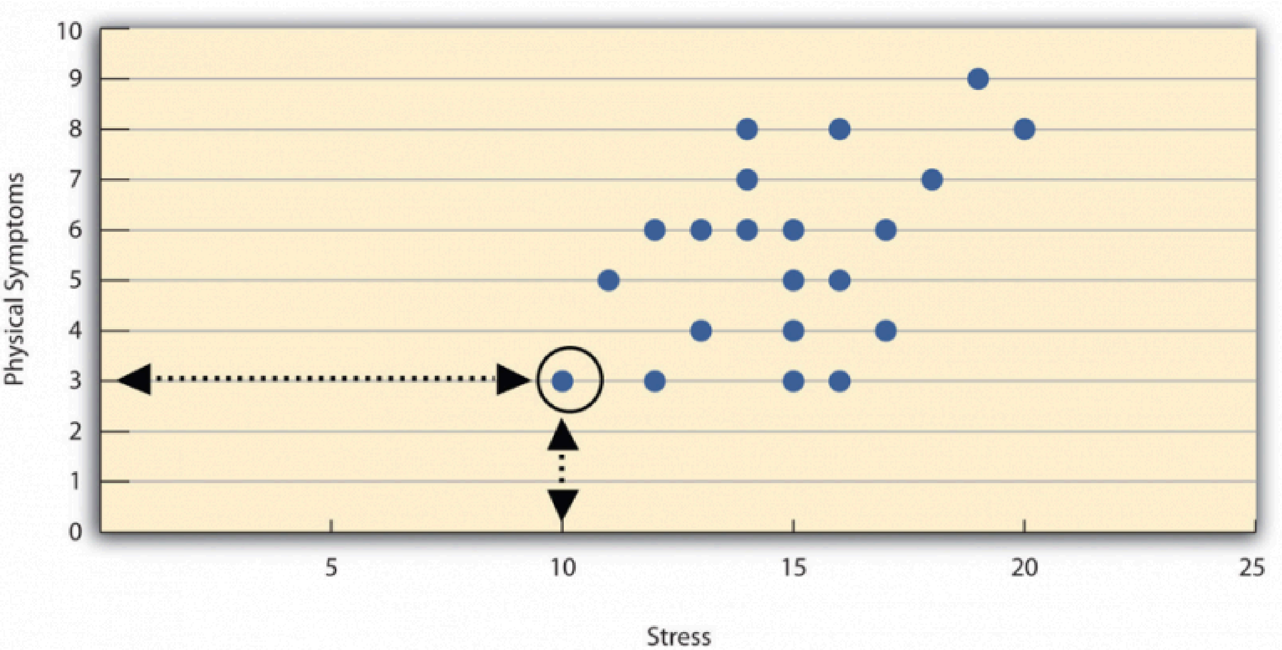


Figure 9.4.2. Scatterplot showing a hypothetical positive relationship between stress and number of physical symptoms. The circled point represents a person whose stress score was 10 and who had three physical symptoms. Pearson's r for these data is +.51 by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

As illustrated in Figure 9.4.3, the strength of this relationship is measured using Pearson's Correlation Coefficient (*r*), which ranges from -1.00 to +1.00. A value of +1.00 represents a perfect positive correlation, while -1.00 represents a perfect negative correlation. A value of 0 indicates no relationship between the two variables. In a scatterplot, a correlation near 0 will appear as a random "cloud" of points, while values closer to ±1.00 show points aligning more closely along a straight line. Correlation coefficients around ±0.10 are considered small, around ±0.30 are considered medium, and around ±0.50 are considered large. It is important to note that the sign (+ or -) indicates the direction of the relationship, not its strength. For example, +0.30 and -0.30 represent relationships of equal strength, but one is positive and the other negative.

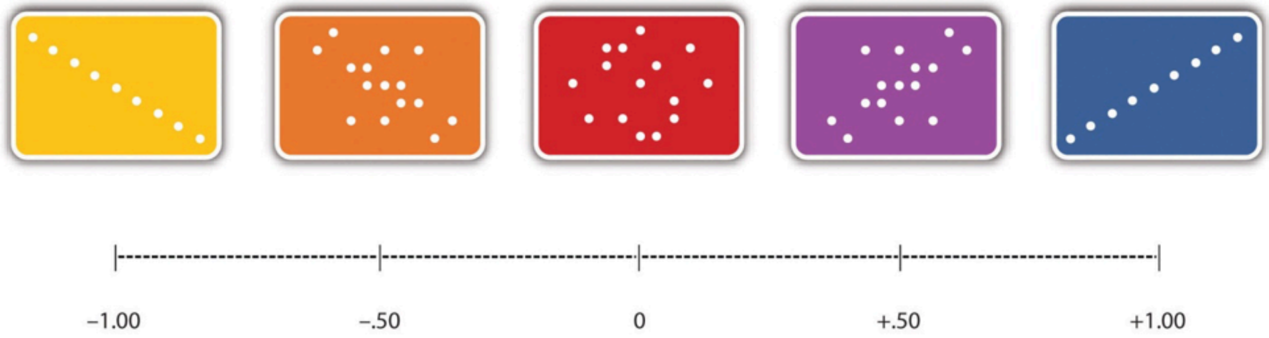


Figure 9.4.3. Range of Pearson's r, from −1.00 (strongest possible negative relationship), through 0 (no relationship), to +1.00 (strongest possible positive relationship) by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

However, there are two scenarios where Pearson's r can be misleading:

Nonlinear Relationships

Pearson's *r* works well for linear relationships, where the data points align roughly along a straight line. However, it is not suitable for nonlinear relationships, where the best-fit line would be a curve. For example, Figure 9.4.4 illustrates the relationship between sleep duration and depression, forming an upside-down U-shape. This hypothetical data suggests that people who sleep too little or too much may show higher depression levels, while those who sleep around eight hours have the lowest depression levels. Even though there is a strong relationship, Pearson's *r* would be close to zero because it does not fit a straight line. This highlights the importance of creating a scatterplot first to visually check if the relationship is linear before relying on Pearson's *r*.

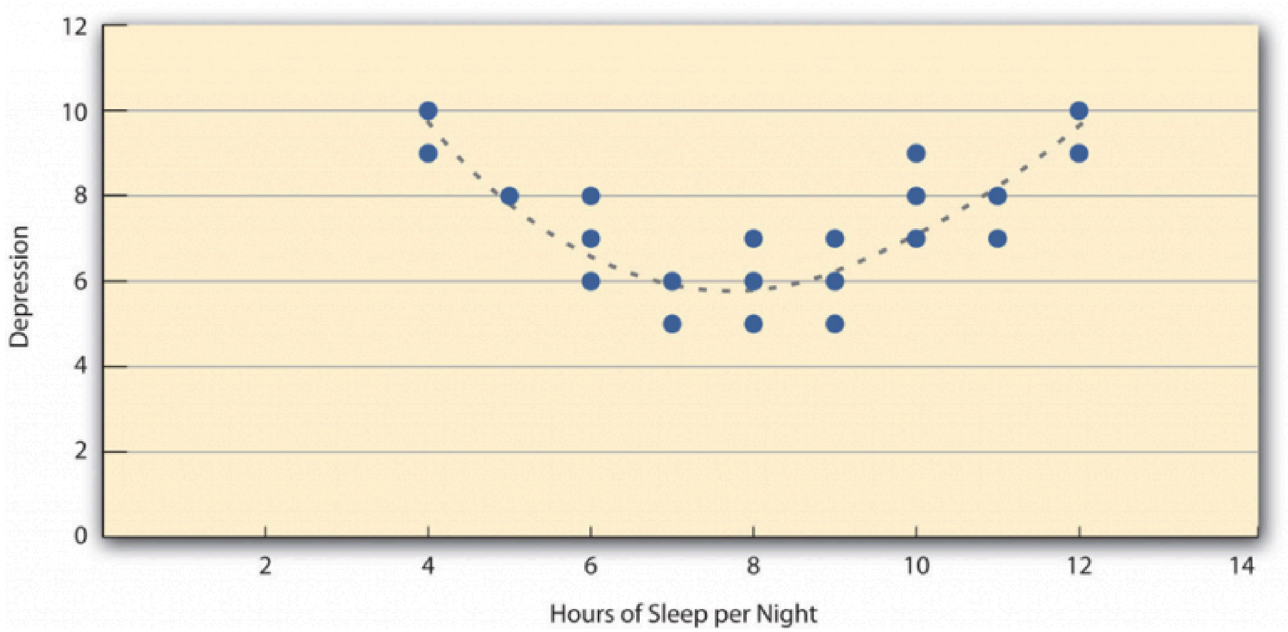


Figure 9.4.4. Hypothetical nonlinear relationship between sleep and depression by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Restriction of Range

Pearson's *r* can also be misleading when there is limited variation in one or both variables. This issue, called restriction of range, occurs when the sample does not represent the full spectrum of values. For instance, imagine a strong negative correlation between age and enjoyment of hip-hop music across a wide range of ages. However, if the sample only includes 18- to 24-year-olds, as shown in the blue box in Figure 9.4.5, the correlation might disappear entirely because this group shows limited variation in music preference. To avoid this, researchers should aim to include a wide range of values for their key variables. If restriction of range happens unintentionally, it is essential to interpret Pearson's *r* cautiously and consider statistical adjustments when necessary.

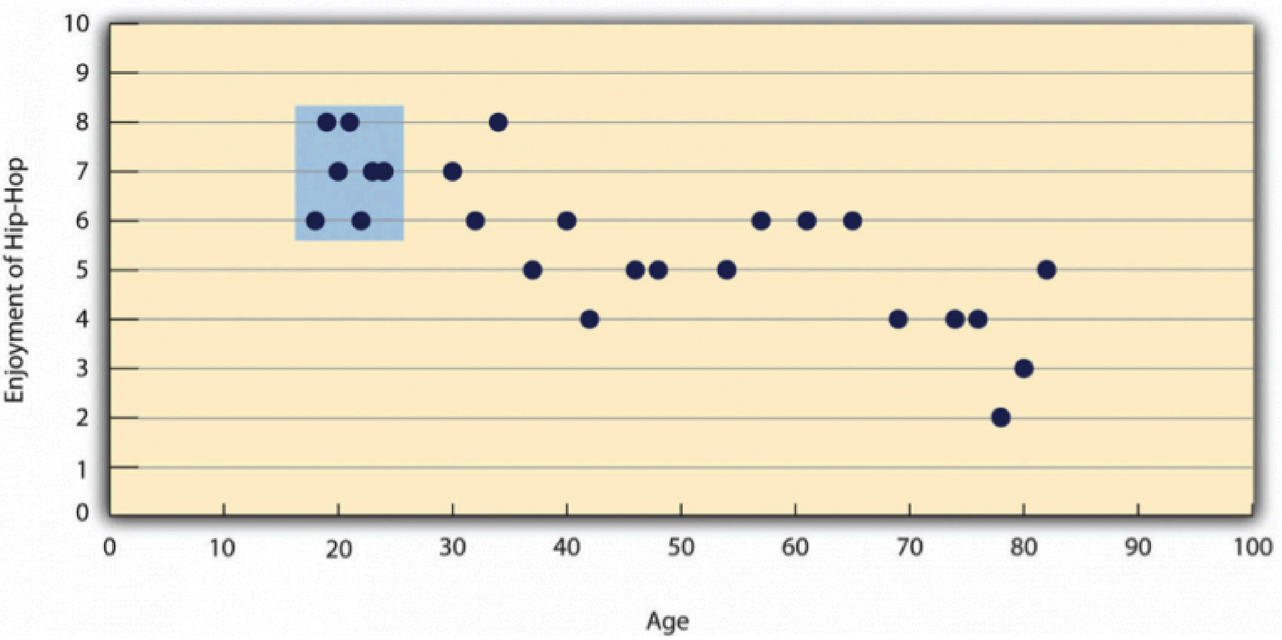


Figure 9.4.5. Hypothetical data showing how a strong overall correlation can appear to be weak when one variable has a restricted range. The overall correlation here is −.77, but the correlation for the 18- to 24-year-olds (in the blue box) is 0 by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Correlation Does Not Imply Causation

You have likely heard the phrase, "Correlation does not imply causation". A humorous example comes from a 2012 study showing a strong positive correlation (Pearson's $r = 0.79$) between a country's per capita chocolate consumption and the number of Nobel Prizes awarded to its citizens. While the data suggest a relationship, it is unlikely that eating more chocolate directly causes people to win Nobel Prizes. Recommending parents to give their children more chocolate in hopes of future Nobel success would not make much sense.

There are two key reasons why correlation does not imply causation:

1. **The Directionality Problem:** When two variables (X and Y) are correlated, it could mean that X causes Y, or it could mean that Y causes X. For example, studies often show that people who exercise regularly tend to be happier. While it is tempting to conclude that exercise causes happiness, the reverse could also be true, where happier people might be more inclined to exercise, perhaps because they have more energy or seek social interaction at the gym.
2. **The Third-Variable Problem:** Sometimes, two variables (X and Y) are correlated not because one causes the other, but because a third variable (Z) affects both. For instance, the correlation between chocolate consumption and Nobel Prize wins could actually stem from a third factor, like geographic or economic factors. European countries, for example, often have high chocolate consumption and also invest significantly in education and scientific research. Similarly, the link between exercise and happiness could be explained by a third variable, such as physical health, where healthier people are both more likely to exercise and to feel happier.

Correlations caused by third variables are often referred to as spurious correlations. These are relationships that appear meaningful but are actually coincidental or driven by an unseen factor.

For more entertaining examples of spurious correlations, you can visit Tyler Vigen's website, where you will find charts showing amusing and unlikely connections between seemingly unrelated variables, such as the one shown in Figure 9.4.6.

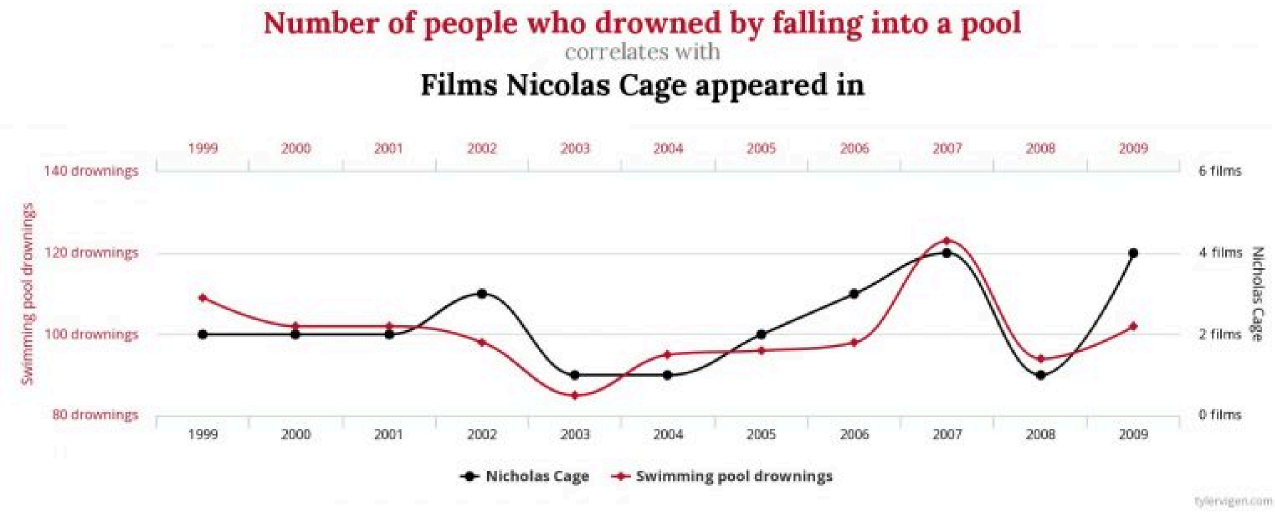


Figure 9.4.6. Example of a spurious correlation. Source: http://tylervigen.com/spurious-correlations is used under a CC-BY 4.0 licence

"Lots of Candy Could Lead to Violence" – a Misleading Claim

While psychologists understand that correlation does not imply causation, many journalists often overlook this principle. A website dedicated to exploring this issue, Correlation or Causation?, highlights numerous media reports where headlines suggest a causal relationship when the evidence only shows a correlation. These misunderstandings usually stem from directionality problems or third-variable problems.

One example is a study showing that children who ate candy every day were more likely to be arrested for violent offences later in life. The headline suggested that candy could "lead to" violence. But is this really true? Could there be other explanations for this statistical relationship? Perhaps children who ate more candy also grew up in environments with less parental supervision or had less balanced diets, both of which could contribute to behavioural issues. A more accurate headline might say, "Study Finds Link Between Daily Candy Consumption and Later Behavioural Problems," instead of implying direct causation.

Addressing the Directionality and Third-Variable Problems

Researchers use specific strategies to address these issues, and the most effective approach is to conduct an experiment. For example, rather than just measuring how much people exercise, a researcher might randomly assign participants to either run on a treadmill for 15 minutes or sit on a couch for 15 minutes.

This minor change in the research design is incredibly powerful. If the participants who exercised show improved moods compared to those who did not, the causal link becomes clearer. This is because:

- The directionality problem is eliminated, since mood cannot influence exercise levels because the researcher controlled who exercised and who did not.
- The third-variable problem is minimised, since factors like physical health or energy levels cannot explain the mood difference since random assignment balanced those variables across the groups.

Assessing Relationships Among Multiple Variables

In correlational research, psychologists often measure multiple variables, whether binary (e.g., yes/no) or continuous (e.g., numerical scores), to understand the statistical relationships among them. For example, Nathan Radcliffe and William Klein conducted a study on middle-aged adults to explore how optimism (measured using the Life Orientation Test) is related to factors associated with heart attack risks. These factors included participants' overall health, their knowledge of heart attack risk factors, and their perceptions of their own heart attack risk. The researchers found that more optimistic individuals tended to exercise more, had lower blood pressure, were more informed about heart attack risk factors, and accurately perceived their risk as lower than that of their peers.

Another study by Ernest Jouriles and colleagues examined the relationship between adolescents' experiences of physical and psychological aggression in relationships and their levels of psychological distress. Since measures of physical aggression often produce skewed results, the researchers converted the data into binary categories (0 = did not occur, 1 = did occur). They did the same for psychological aggression and then analysed the relationships between these variables. The results revealed that adolescents who experienced physical aggression were also moderately likely to experience psychological aggression, and those exposed to psychological aggression reported higher psychological distress.

This approach is also frequently used to validate new psychological measures. For instance, when John Cacioppo and Richard Petty developed the Need for Cognition Scale (a tool measuring how much people enjoy and value thinking), they tested it on a large sample of college students. They also measured intelligence, socially desirable responding (the tendency to give what one believes is the "right" or socially acceptable answer), and dogmatism. As shown in Table 9.4.1, the findings were presented in a correlation matrix showing statistical relationships (Pearson's *r*) between each pair of variables. For example, the correlation between need for cognition and intelligence was +.39, while the correlation between intelligence and socially desirable responding was +.02.

The overall pattern of correlations supported the researchers' expectations about how need for cognition scores should relate to these other psychological constructs.

Table 9.4.1. Correlation matrix showing correlations among the need for cognition and three other variables based on research by Cacioppo and Petty (1982)

	Need for cognition	Intelligence	Social desirability	Dogmatism
Need for cognition	—			
Intelligence	+.39	—		
Social desirability	+.08	+.02	—	
Dogmatism	−.27	−.23	+.03	—

Factor Analysis

Factor analysis is a statistical method used to identify patterns among a large set of related variables. It works by grouping these variables into smaller clusters, where variables within each cluster are strongly correlated with one another but show weaker correlations with variables in other clusters. Each cluster represents an underlying concept or "factor" that captures the shared characteristics of the variables within it.

For example, when people are tested on various mental tasks, factor analysis often reveals two primary factors: one associated with mathematical intelligence (tasks like arithmetic, quantitative estimation, and spatial reasoning) and the other with verbal intelligence (tasks like grammar, reading comprehension, and vocabulary). Similarly, the Big Five personality traits, such as extraversion, openness, and conscientiousness, were identified through factor analysis by grouping specific traits that tend to co-occur.

Another interesting application of factor analysis comes from research by Peter Rentfrow and Samuel Gosling. They asked over 1,700 university students to rate their preferences across 14 music genres. Factor analysis revealed four distinct clusters or dimensions of musical preference:

1. **Reflective and Complex:** Blues, jazz, classical, and folk music
2. **Intense and Rebellious:** Rock, alternative, and heavy metal music
3. **Upbeat and Conventional:** Country, soundtrack, religious, and pop music
4. **Energetic and Rhythmic:** Rap/hip-hop, soul/funk, and electronica music

These clusters suggest that musical preferences are not random but follow meaningful patterns that reflect broader tastes.

However, it is important to note that factors are not categories. Factor analysis does not classify people into rigid groups. For instance, someone high in extraversion might also score high or low in conscientiousness. Similarly, someone who enjoys reflective and complex music might also like intense and rebellious music. Additionally, factors reveal structure, not meaning. Factor analysis identifies patterns, but it is up to researchers to interpret and label these factors meaningfully. For example, researchers have suggested that the Big Five personality factors arise from different genetic influences.

Exploring Causal Relationships in Correlational Research

While it is true that correlation does not imply causation, correlational research can still offer valuable insights into potential causal relationships between variables. Although it cannot definitively prove that one variable causes another, it can help rule out alternative explanations using a technique called partial correlation. Instead of controlling third variables through random assignment or holding them constant (as in experiments), researchers measure these third variables and include them in statistical analyses.

Imagine a researcher investigating whether watching violent television shows leads to aggressive behaviour. She suspects that socioeconomic status (SES) might be a third variable influencing this relationship. To address this, she conducts a study measuring three factors: the amount of violent television participants watch, their aggressive behaviours, and their SES.

First, she calculates the correlation between violent TV viewing and aggression and finds a moderate positive correlation of +0.35. Next, she uses partial correlation to control for SES. This analysis isolates the relationship between violent TV viewing and aggression, removing the influence of SES on both variables.

- If the partial correlation remains high (e.g., +0.34), this suggests that the relationship between violent TV viewing and aggression exists independently of SES. SES is not a driving third variable.
- If the correlation drops significantly (e.g., +0.03), this suggests that SES is likely the key factor driving the relationship.
- If the correlation decreases but remains moderate (e.g., +0.20), this indicates that SES explains part of the relationship, but not all of it.

While partial correlation is a powerful statistical tool for addressing third-variable problems, it has its limitations. It does not resolve the directionality problem (we still don't know if watching violent TV causes aggression or if aggressive individuals are drawn to violent TV). Additionally, there could be other unmeasured third variables influencing the relationship that were not accounted for in the analysis.

Understanding Regression in Research

Once a relationship between two variables is identified, researchers can use this relationship to predict one variable based on another. For example, if we know there is a correlation between IQ scores and GPA, we can use someone's IQ score to predict their GPA. While correlation coefficients describe the strength and direction of a relationship, regression analysis goes a step further by providing a statistical model to make predictions.

In a regression analysis, the variable used to make predictions is called the predictor variable, and the variable being predicted is called the outcome variable (or criterion variable). The basic formula for regression looks like this:

$Y = �_1�_1$

- Y represents the predicted score on the outcome variable.
- b_1 is the regression weight (slope), which shows how much the outcome variable changes for each one-unit increase in the predictor variable.
- X_1 represents the person's score on the predictor variable.

To predict someone's score on the outcome variable (Y), you simply multiply their predictor score (X_1) by the regression weight (b_1).

Simple vs. Multiple Regression

While simple regression involves predicting an outcome using one predictor variable, multiple regression uses several predictor variables (X_1, X_2, X_3,…X_i) to predict an outcome variable (Y). The general formula for multiple regression looks like this:

$� = �_1�_1 + �_2�_2 + �_3�_3 + ... + �_i�_i$

- Each *b* (e.g., b_1, b_2) represents how much each predictor variable contributes to predicting the outcome variable.
- The regression weights show how much the outcome variable changes with a one-unit increase in each predictor variable, assuming all other variables remain constant.

Why Use Multiple Regression?

Multiple regression is valuable because it can show how each predictor variable contributes to an outcome variable while statistically controlling for other predictor variables.

For example, imagine a researcher wants to understand how income and health relate to happiness. These two predictors (income and health) are related to each other, making it difficult to tell if one directly influences happiness or if the relationship is indirect.

- If wealthier people are happier, is it because they are healthier?
- If healthier people are happier, is it because they tend to have more money?

Using multiple regression, researchers can statistically control for one variable while examining the other. This means they can isolate the effect of income on happiness, independent of health, and vice versa.

Research using multiple regression has shown that both income and health contribute only slightly to happiness, except in extreme cases of poverty or severe illness (Diener, 2000).

However, while regression analysis is powerful for identifying patterns and predicting outcomes, it cannot definitively establish causation. At best, it reveals patterns of relationships that are consistent with some causal explanations and inconsistent with others. To establish causality, experimental designs remain the gold standard.

References

Bushman, B. J. & Huesmann, L. R. (2011). Effects of violent media on aggression. In D. G. Singer & J. L. Singer (Eds.), *Handbook of children and the media* (2nd ed., pp. 231-248). Sage.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.

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

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

9.5. QUALITATIVE RESEARCH

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

While most psychological research is quantitative, focusing on numerical data, statistical analysis, and broad generalisations, there is an equally important approach known as qualitative research. This method originated in anthropology and sociology but is now widely used in psychology to gain deeper insight into human experiences.

Quantitative researchers typically start with a specific research question or hypothesis, gather numerical data from a large group of participants, and use statistical techniques to draw conclusions that can be generalised to a broader population. In contrast, qualitative researchers begin with a broader research question, collect large amounts of detailed, non-numerical data from a smaller group of participants, or other types of naturalistic data, such as observations, photographs, and policy documents. Qualitative researchers analyse them using a variety of analytical strategies such as thematic analysis, grounded theory, critical discourse analysis, or interpretative phenomenological analysis. Their goal is less about drawing universal conclusions and more about gaining a deeper understanding of participants' individual experiences and socio-cultural contexts.

For example, in a study by Per Lindqvist and colleagues (2008), researchers aimed to understand how families of teenage suicide victims cope with their loss. Instead of asking a precise question like, "What percentage of family members join support groups?", they focused on capturing the range of emotional experiences these families faced.

To gather their data, the researchers conducted unstructured interviews with families of 10 teenage suicide victims in rural Sweden. These interviews began with an open-ended request for participants to share their thoughts about their loved one and ended with an invitation to add anything else they felt was important.

One significant theme that emerged was that even when families' daily lives returned to a sense of "normalcy", they continued to struggle with the unanswered question of why the suicide happened. This emotional struggle was particularly intense in cases where the suicide was unexpected.

The Purpose of Qualitative Research

Quantitative research excels at answering specific research questions with precision and drawing general conclusions about human behaviour. For instance, it has been shown that people tend to obey authority figures and that female undergraduate students are not significantly more talkative than male students. However, while quantitative research is powerful for answering clear, predefined questions, it falls short in generating new and intriguing research questions, providing detailed descriptions of behaviour in specific contexts, and capturing the personal experience of being part of a particular group or situation.

This is where qualitative research shines. It is especially valuable for generating fresh research questions and hypotheses that can later be tested quantitatively. For example, the study by Lindqvist and colleagues hinted at a potential relationship between how unexpected a suicide is and how deeply families struggle to understand the reasons behind it. This insight emerged directly from listening to families' stories in an open-ended interview setting. Without these intimate conversations, such a question might never have been considered.

Qualitative research also excels at creating rich, detailed descriptions of human behaviour in real-world settings. Researchers often refer to this level of detail as "thick description" (Geertz, 1973). Unlike quantitative studies, qualitative research can capture the "lived experience" of participants. In Lindqvist's study, families frequently offered to show the interviewer their loved one's bedroom or the site of the suicide. This small yet powerful detail highlighted the emotional significance of these spaces, a discovery unlikely to emerge from a quantitative survey. Some of the differences between qualitative and quantitative research are provided in Table 9.5.1.

Table 9.5.1. Some contrasts between qualitative and quantitative research

Qualitative	Quantitative
1. In-depth information about relatively few people	1. Less depth of information with larger samples
2. Conclusions are based on interpretations drawn by the investigator	2. Conclusions are based on statistical analyses
3. Global and exploratory	3. Specific and focused
4. To describe a situation, to identify previously unknown processes: what, why and how	4. To measure magnitude, pervasiveness, and central patterns of association

Data Collection and Analysis in Qualitative Research

Qualitative research employs a diverse range of data collection methods, including naturalistic observation, participant observation, archival data analysis, and artwork analysis, among others. However, one of the most common methods in psychology is conducting interviews.

Interviews in qualitative research can vary in structure:

- Unstructured interviews involve open-ended prompts or general questions, allowing participants to guide the conversation.
- Structured interviews follow a strict script, with little room for deviation.
- Semi-structured interviews fall between these two approaches. Researchers prepare a set of key questions but have the flexibility to ask follow-up questions based on participants' responses.

Semi-structured interviews are widely used because they strike a balance between consistency and adaptability. These interviews are often lengthy and detailed, but are usually conducted with a small group of participants. For example, in Lindqvist and colleagues' study on families of suicide victims, the researchers used unstructured interviews to allow participants to share their experiences at their own pace. This approach ensured that families had control over how much they disclosed about such a sensitive topic. However, conducting interviews can be time-consuming, and interviewers must possess good listening and probing skills to ensure that rich, detailed insights are gathered from participants.

Another popular qualitative method is the use of focus groups. Focus groups are typically used to explore community norms, cultural values, and group opinions. In focus groups, a small group of people discusses a specific topic or issue, usually guided by a facilitator. This group dynamic can encourage richer discussions and reveal insights that might not emerge in one-on-one interviews. Focus groups are widely used in business and industry to understand consumer preferences and behaviours. However, researchers must remain aware of group dynamics within focus groups that can influence the data. For example:

- participants may give socially desirable answers to be liked or accepted by others
- dominant personalities, such as highly extraverted participants, might control the discussion, overshadowing quieter group members.

Data Analysis in Qualitative Research

While qualitative and quantitative research differ in various ways, like how data are collected or the specificity of the research question, the most significant difference lies in how the data are analysed. To illustrate this, imagine a research team conducting unstructured interviews with individuals recovering from alcohol use disorder. Their goal is to understand the role of religious faith in recovery. Initially, this sounds like qualitative research. However, if the team later codes the data based on how frequently participants mention God or a "higher power" and then uses statistical analysis to link these mentions to recovery success, the study shifts into the quantitative research category. This example highlights how data are analysed defines whether research is qualitative or quantitative, rather than how it was collected.

Thematic Analysis

Just as there are many ways to collect qualitative data (e.g., interviews, focus groups, observations), there are also many ways to analyse it. One widely used qualitative analytical method in Psychology is Thematic Analysis. Broadly, Thematic Analysis is used to identify, analyse and report themes within a qualitative dataset. However, there are many varying approaches in Thematic Analysis that differ in the analytical procedures as well as underlying research values or philosophy. Some commonly used approaches include Reflexive Thematic Analysis (Braun & Clarke, 2008, 2021), Applied Thematic Analysis (Guest et al., 2011), and Framework Analysis (Ritchie et al., 2014). The analysis of data using Thematic Analysis typically goes through four main stages:

1. **Familiarising with the qualitative dataset:** This initial step entails researchers reading and re-reading the data (e.g., transcripts, field notes) to become deeply familiar with participants' experiences and perspectives.
2. **Identifying Repeated Ideas:** Researchers look for patterns or recurring ideas within the data through a systematic coding process.
3. **Organising Ideas into Themes:** These repeated ideas are grouped iteratively into broader, meaningful themes or categories.
4. **Creating a Theoretical Narrative:** The final step is crafting an interpretation of the data, often supported by direct participant quotes to bring the analysis to life.

Grounded Theory in Practice

Another popular qualitative research methodology is Grounded Theory, which was first developed by Glaser and Strauss (1967). This approach does not start with a pre-existing theory. Instead, this methodology aims to develop an explanatory theory of psychological or social processes that is "grounded in" the data. Consider a study by Laura Abrams and Laura Curran (2009), which explored the experiences of postpartum depression among low-income mothers. They conducted unstructured interviews with 19 participants and identified five major themes from the data, each made up of smaller, repeating ideas. For instance:

Table 9.5.2. Categories and repeating ideas in a study of postpartum depression among low-income mothers. Based on research by Abrams and Curran (2009)

Category	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."

In their findings in Table 9.5.2, Abrams and Curran emphasised that the participants' postpartum symptoms were not just abstract "affective disorders". Instead, their symptoms were deeply tied to the daily struggles of raising children alone, often in challenging and unstable circumstances.

A participant codenamed "Destiny" shared this perspective:

> 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.

This example illustrates how Grounded Theory enables researchers to develop a rich, nuanced understanding of participants' experiences, capturing both their struggles and the contexts that shape them.

The Quantitative-Qualitative Debate

Quantitative and qualitative research approaches in psychology and related fields have often been viewed as opposing methods. Each has its strengths and limitations, which have led to ongoing debates about their value and effectiveness.

Some quantitative researchers argue that qualitative methods lack objectivity and are hard to evaluate in terms of reliability and validity. They also point out that qualitative findings are often difficult to generalise to broader populations or other contexts. On the other hand, some qualitative researchers criticise quantitative methods for oversimplifying the richness of human behaviour and focusing only on easily measurable variables.

However, both sides are aware of these criticisms. Qualitative researchers have developed frameworks to address concerns about objectivity, reliability, validity, and generalisability (though these frameworks are beyond our current scope). Similarly, quantitative researchers recognise that human behaviour cannot always be reduced to a few variables and statistical relationships. For them, simplification is a strategic tool for uncovering general principles of behaviour rather than a belief that behaviour itself is simple.

A Middle Ground: Mixed-Methods Research

Today, many researchers agree that quantitative and qualitative approaches can complement each other. This integrated approach is known as mixed-methods research (Todd et al., 2004). In fact, studies like those conducted by Lindqvist et al. and Abrams & Curran combined both approaches effectively.

Two Common Mixed-Methods Approaches

Hypothesis Generation and Testing:

- Qualitative research can generate hypotheses by exploring complex behaviours, experiences, or patterns in small samples.
- Quantitative research can then test those hypotheses using larger, more controlled samples.
- For example, if a qualitative study suggests that families struggle more with unresolved questions after an unexpected suicide, a quantitative study could measure this relationship in a larger sample to confirm or refine the hypothesis.

Triangulation:

- This approach uses both qualitative and quantitative methods simultaneously to study the same question from different angles.
- If both methods produce similar results, they reinforce and enrich each other.
- If the results diverge, it raises an interesting follow-up question: Why did they differ, and how can this be explained?

An Example of Mixed-Methods Research

A study by Trenor et al. (2008) investigated the experiences of female engineering students at a university. In the first phase, students completed a quantitative survey where they rated perceptions such as their sense of belonging. Statistical analysis showed no significant differences across ethnic groups in their ratings. At first glance, this result could suggest that ethnicity does not influence students' sense of belonging.

However, in the second phase, researchers conducted qualitative interviews with the same students. During these interviews, many minority students reported that cultural diversity on campus enhanced their sense of belonging. Without the qualitative insights, the quantitative results might have been misinterpreted, leading to an incomplete understanding of the data.

This example demonstrates how quantitative and qualitative methods complement each other, offering a more complete picture of human behaviour. While qualitative research excels at identifying patterns, behaviours, and unique experiences, quantitative research is better suited for testing hypotheses, identifying statistical relationships, and uncovering broader trends.

Some researchers suggest that qualitative research helps identify phenomena, while quantitative research helps explain mechanisms. However, Bryman (2016) argues that we should move past this artificial divide and recognise that both approaches are valuable tools for investigating complex questions about human behaviour.

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/10.1177/036168430903300309

Braun, V., & Clarke, V. (2008). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage Publications.

Bryman, A. (2012). *Social research methods* (4th ed). Oxford University Press.

Geertz, C. (1973). *The interpretation of cultures*. Basic Books.

Glaser, B. G., Strauss, A. L., & ProQuest. (1999). *The discovery of grounded theory strategies for qualitative research*. Aldine de Gruyter.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Sage Publications.

Ritchie, J., Lewis, J., McNaughton Nicholls, C., & Ormston, R. (Eds.). (2014). *Qualitative research practice: A guide for social science students and researchers* (2nd ed.). Sage.

Todd, Z. (2004). *Mixing methods in psychology: The integration of qualitative and quantitative methods in theory and practice*. Psychology Press. https://doi.org/10.4324/9780203645727

Trenor, J.M., Yu, S.L., Waight, C.L., Zerda, K.S. and 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*, (449-465). https://doi.org/10.1002/j.2168-9830.2008.tb00992.x

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

9.6. SURVEY RESEARCH

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Survey research is a method used in both quantitative and qualitative studies, characterised by two key features.

First, self-reports are used to measure variables. In quantitative research, researchers typically gather information through questionnaires, asking participants, often referred to as respondents, to share their thoughts, feelings, and behaviours directly. Such questionnaires may include some open-ended questions. In fully qualitative surveys, participants respond in their own words to a series of pre-determined open-ended questions. These responses are subsequently analysed to provide in-depth insights into participants' framework of meanings, including how they understand and interpret their social world.

Second, sampling plays a crucial role. Survey researchers prioritise large random samples because they provide the most accurate representation of the larger population. In fact, survey research is one of the few approaches in psychology where random sampling is a standard practice.

Beyond these two defining features, surveys are highly flexible. They can vary in length (short or long) and format (in-person, telephone, mail, or online). Surveys can cover a wide range of topics, including voting preferences, consumer habits, social attitudes, health issues, or anything else people can meaningfully respond to. While survey data are often analysed using statistical methods, some research questions are better suited for qualitative analysis.

Non-Experimental and Experimental Survey Research

Most survey research is non-experimental and focuses on two primary goals:

- describing single variables, such as identifying the percentage of voters supporting a candidate or the prevalence of a mental health condition
- assessing relationships between variables, like studying the connection between income and health.

However, surveys can also be used in experimental research. For example, in the aftermath of the September

2001 terrorist attacks, Jennifer Lerner and her colleagues surveyed nearly 2,000 American teens and adults to examine their emotional reactions and risk perceptions. They found that participants primed to feel anger perceived less risk, while those primed to feel fear perceived greater risk, highlighting how specific emotions influence risk assessment. In their study, they used self-report measures and a large national sample, which aligns with typical survey research. However, they also manipulated an independent variable (e.g., inducing feelings of anger or fear) to observe its effect on a dependent variable (e.g., risk judgements), making it an experimental study as well.

History and Uses of Survey Research

Survey research traces its origins to early 20th-century social surveys in England and the United States, where researchers aimed to document social issues such as poverty (Converse, 1987). By the 1930s, the U.S. government began using surveys to understand economic and social conditions, leading to advancements in sampling techniques. Around the same time, market researchers transitioned into election polling, refining survey methods to predict public opinion more accurately.

A pivotal moment occurred during the 1936 U.S. presidential election between Alf Landon and Franklin Roosevelt. The magazine Literary Digest predicted a Landon victory based on a massive mail survey. In contrast, George Gallup, using smaller but scientifically selected samples, predicted a Roosevelt landslide. Gallup's prediction was correct, proving the power of methodologically sound surveys. This success laid the foundation for the first national election survey in 1948 by the Survey Research Centre at the University of Michigan, now known as the American National Election Studies.

Survey research soon became a cornerstone in fields such as political science, sociology, and public health. In the 1930s, psychologists made significant contributions to questionnaire design, including the widely used Likert scale. Surveys also played a key role in the social psychological study of attitudes, stereotypes, and prejudice, encouraging the use of larger and more diverse samples beyond the convenience of university students.

Today, survey research remains essential in psychology. Large-scale surveys, like the National Comorbidity Survey, have been instrumental in estimating the prevalence of mental disorders and exploring their relationships with other factors. For example, nearly 10,000 adults were interviewed in their homes in 2002–2003, revealing valuable insights into the lifetime prevalence of disorders such as generalised anxiety, depression, and substance abuse (see Table 9.6.1).

Table 9.6.1. Some lifetime prevalence results from the National Comorbidity Survey

Lifetime prevalence*			
Disorder	**Total**	**Female**	**Male**
Generalised anxiety disorder	5.7	7.1	4.2
Obsessive-compulsive disorder	2.3	3.1	1.6
Major depressive disorder	16.9	20.2	13.2
Bipolar disorder	4.4	4.5	4.3
Alcohol abuse	13.2	7.5	19.6
Drug abuse	8.0	4.8	11.6

***The lifetime prevalence of a disorder is the percentage of people in the population who develop that disorder at any time in their lives.**

Survey research is also flexible enough to be integrated into experimental studies to test causal hypotheses. Large and diverse survey samples can complement laboratory studies, offering a broader perspective on psychological phenomena.

Constructing Surveys

The core of any survey research project is the survey itself. While coming up with interesting questions may seem simple, creating a well-designed survey is far from easy. The way people respond can be unintentionally influenced by factors such as how questions are worded, the order in which they appear, the response options provided, and other subtle details.

At best, these factors add random noise to the data, making it harder to identify clear patterns. At worst, they create systematic biases that can lead to misleading conclusions.

To ensure that survey responses are as accurate and trustworthy as possible, researchers must carefully design their surveys to minimise these unintended effects.

Survey Responding as a Psychological Process

Before diving into the specific principles of designing a survey, it is helpful to understand that answering survey questions is a psychological process. Respondents do not just provide answers. Instead, they go through a series of mental steps to interpret questions, retrieve relevant information from memory, make judgements, and then choose a response.

A Cognitive Model of Survey Responding

When people respond to a survey question, they go through several mental steps, as illustrated in the cognitive model developed by Sudman et al. (1996). Figure 9.6.1 lists the steps involved in the cognitive model of survey responding.

Figure 9.6.1. Model of the cognitive processes involved in responding to a survey item by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Take this example question:

How many alcoholic drinks do you consume in a typical day?

- a lot more than average
- somewhat more than average
- average
- somewhat fewer than average
- a lot fewer than average.

At first glance, the question seems simple, but it actually presents multiple challenges.

Step 1: Interpreting the Question

Respondents must first understand the question. They need to decide if "alcoholic drinks" include beer, wine and hard liquor, or just hard liquor. They must also decide if "a typical day" refers to a weekday, a weekend day, or both. Although research by Chang and Krosnick (2003) suggests asking about "typical behaviour" is more valid than asking about "past behaviour", the interpretation of "typical day" may still vary depending on the respondent's habits.

Step 2: Retrieving Information

Next, respondents must recall relevant memories. Should they think about specific recent drinking occasions? Should they carefully count the number of drinks they had last week and divide by seven? Or should they rely on general beliefs they hold about themselves, like "I'm not much of a drinker"?

Step 3: Forming a Tentative Judgement

Once they have gathered information, respondents must calculate an estimate. For example, they might average their drinks from the past week to form an answer.

Step 4: Choosing a Response

Now they must match their tentative judgement to the response options provided. This step brings additional challenges. For instance, what does "average" mean? How much is "somewhat more than average"? These terms are subjective and open to interpretation.

Step 5: Editing the Response

Finally, respondents may adjust their answer before reporting it. For example, someone who drinks heavily might hesitate to admit they drink "a lot more than average" and instead choose the less extreme option, "somewhat more than average", to avoid judgment.

What seems like a simple question, "How much do you drink?", actually involves a complex series of cognitive steps. Each step introduces potential sources of error, highlighting the importance of careful survey design to ensure clarity and reduce ambiguity.

Context Effects on Survey Responses

Survey responses can be unintentionally influenced by factors unrelated to the actual content of the questions. These influences, known as context effects, arise from the way questions are presented or the options provided (Schwarz & Strack, 1991).

One common context effect is the item-order effect, where the order of questions impacts how respondents answer. For example, in a study by Fritz Strack and colleagues (1988), college students were asked about their overall life satisfaction and their dating frequency. When the life satisfaction question was asked first, the correlation between the two answers was weak (−0.12). However, when the dating question came first, the correlation became strong (+0.66). This shift happened because answering the dating question first made dating frequency more prominent in the respondents' minds, influencing how they rated their overall life satisfaction.

Another context effect comes from the response options provided (Schwarz, 1999). For example, when people are asked how often they feel "really irritated" and given options ranging from "less than once a year" to "more than once a month", they tend to interpret the question as referring to major irritations and report being irritated less frequently. However, if the options range from "less than once a day" to "several times a month," they focus on minor irritations and report being irritated more frequently.

Additionally, respondents often assume that middle response options represent what is normal or typical. For example, when asked about how much television they watch, people are more likely to report higher viewing times if the middle option is set at 4 hours rather than 2 hours.

To reduce these context effects, researchers can rotate or randomise the order of questions and response options, especially when no natural order exists. Online survey tools often allow for counterbalancing or randomisation, which helps minimise order effects. This practice is particularly important in contexts like political polls, where being listed first on a ballot has been shown to give candidates a 2.5% boost among undecided voters simply due to their position on the list (Miller & Krosnick, 1998).

Writing Survey Items

Types of Survey Items

Survey items generally fall into two categories: open-ended and closed-ended. Each type serves a specific purpose and has its strengths and limitations.

Open-Ended Items

Open-ended questions allow participants to respond freely without being constrained by preset options. Examples include:

- "What is the most important thing to teach children to prepare them for life?"
- "Please describe a time when you were discriminated against because of your age."
- "Is there anything else you would like to tell us about?"

Open-ended questions are particularly useful when researchers are uncertain about the range of possible responses or want to avoid influencing participants' answers. These questions are often employed in the early stages of research or when exploring qualitative insights.

While open-ended questions are relatively easy to write, they come with challenges:

- they require more time and effort from participants
- they are harder to analyse because responses must be transcribed, coded, and analysed qualitatively
- participants are more likely to skip open-ended questions because they require longer answers.

Open-ended items are most effective when the researchers are exploring unknown variables or when the data can later be categorised for analysis.

Closed-Ended Items

Closed-ended questions provide participants with a predefined set of response options. Examples include:

- "How old are you?"
    - Under 18
    - 18 to 34
    - 35 to 49

    - 50 to 70
    - Over 70
- "On a scale of 0 (no pain at all) to 10 (worst pain ever experienced), how much pain are you in right now?"
- "Have you ever in your adult life been depressed for a period of 2 weeks or more?"
    - Yes
    - No.

Closed-ended questions are useful when researchers have a clear understanding of possible responses. They are typically used for quantitative analysis and are ideal for measuring specific variables or constructs, such as agreement levels, risk perceptions, or behavioural frequency.

While closed-ended questions are more difficult to write (due to the need for appropriate response options), they are:

- faster and easier for participants to answer
- simpler to analyse, as responses can be easily converted into numerical data and entered into statistical software.

Response Options in Closed-Ended Items

All closed-ended questions include response options designed to match the type of variable being measured. For categorical variables, such as gender or political party preference, participants select one or more applicable categories from a list. For quantitative variables, researchers commonly use rating scales, which present an ordered set of responses for participants to choose from.

Rating scales typically range from three to eleven points, with five-point and seven-point scales being the most popular. Five-point scales are best suited for unipolar questions, where only one dimension is being measured, such as frequency (Never, Rarely, Sometimes, Often, Always). In contrast, seven-point scales are ideal for bipolar questions, where responses fall along a spectrum with two opposing ends, such as liking (Like very much to Dislike very much).

To improve the accuracy of responses on bipolar scales, researchers often use a branching technique. For example, if asking about preferences for ice cream, the first question might be, "Do you generally like or dislike ice cream?" Based on the participant's response, they would then be presented with the corresponding range of options from the seven-point scale. This approach improves both the reliability and validity of the responses, as demonstrated by research from Krosnick and Berent (1993).

When presenting rating scales, it is generally better to use verbal labels rather than numerical ones during data collection, though numerical values are still applied during analysis. Researchers should also avoid overly specific or partial labels, as these can confuse participants.

Another type of response option is the visual-analogue scale, such as the one shown in Figure 9.6.2, where participants mark a point along a horizontal line to indicate their response. These scales are particularly effective when paired with meaningful graphics that help clarify the intended meaning of the response options.

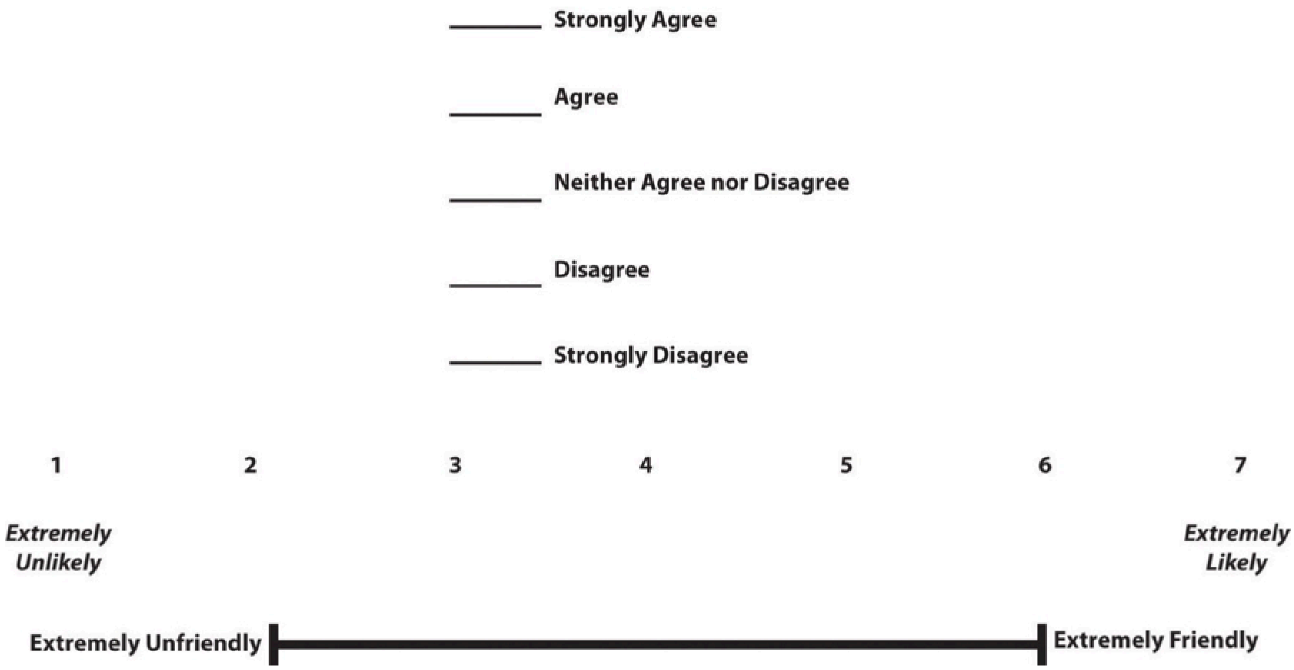


Figure 9.6.2. Example rating scales for closed-ended questionnaire items by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

What is a Likert Scale?

A Likert scale is a specific type of rating scale commonly used in psychological research to measure people's attitudes toward a person, group, or idea. Although the term is often misused to describe any rating scale (e.g., a 0-to-10 life satisfaction scale), it has a precise definition.

In the 1930s, researcher Rensis Likert (pronounced LICK-ert) developed this method to quantify attitudes. Participants are presented with a series of statements, both positive and negative, about a particular subject. They then indicate their level of agreement or disagreement with each statement using a 5-point scale:

- strongly agree
- agree
- neither agree nor disagree
- disagree
- strongly disagree.

Each response is assigned a numerical value, and these values are summed across all items to produce an overall score representing the individual's attitude. For negatively worded statements, researchers use reverse coding to ensure consistency in scoring across all items.

It is important to note that a true Likert scale specifically measures attitudes using multiple statements rated on this 5-point agreement scale. If you are using a single-item rating scale or measuring something other than attitudes (e.g., frequency or satisfaction), it is more accurate to simply call it a rating scale rather than a Likert scale.

Writing Effective Survey Items

Creating effective survey questions is key to minimising confusion and maximising the reliability and accuracy of participants' responses. The BRUSO model (Peterson, 2000) offers a helpful guideline for crafting strong survey items. **BRUSO** stands for **B**rief, **R**elevant, **U**nambiguous, **S**pecific, and **O**bjective.

Brief items are short and direct, avoiding unnecessary words or overly technical language. This makes them easier for participants to understand and quicker to answer. For example, instead of asking, "Are you now or have you ever been the possessor of a firearm?", a better version would be "Have you ever owned a gun?"

Relevant items focus only on information directly related to the research question. Asking about a participant's sexual orientation, marital status, or income should only be included if it is clearly necessary for the study. Irrelevant questions not only waste time but can also irritate respondents.

Unambiguous items leave no room for multiple interpretations. A question like "Are you a gun person?" is unclear, as respondents may have different interpretations of what being a "gun person" means. A clearer version would be "Do you currently own a gun?"

Specific items focus on a single, clear idea. Avoid double-barreled questions, which ask about two separate issues but allow only one answer. For example, "How much have you read about the new gun control measure and sales tax?" should be split into two separate questions about gun control and sales tax.

Objective items should not reveal the researcher's bias or guide respondents toward a particular answer. Instead of asking, "How much do you support the new gun control measure?", a more neutral option would be "What is your view of the new gun control measure?"

Response Options for Closed-Ended Items

For categorical variables, response options should be mutually exclusive and exhaustive.

- Mutually exclusive means the categories do not overlap. For example, "Protestant" and "Catholic" are mutually exclusive, but "Christian" and "Catholic" are not.
- Exhaustive means all possible answers are covered. If including every option is not feasible, an "Other" category with space for respondents to clarify their response can be used.

If respondents can belong to more than one category (e.g., race or ethnicity), they should be instructed to "Select all that apply".

For rating scales, five or seven response options usually offer enough precision without overwhelming respondents. For more familiar dimensions, such as attractiveness, pain, or likelihood, a 0-to-10 scale can also work effectively.

Regardless of the number of response options, the scale should be balanced around a neutral midpoint. For example:

- **Unbalanced Scale:** Unlikely | Somewhat Likely | Likely | Very Likely | Extremely Likely
- **Balanced Scale:** Extremely Unlikely | Somewhat Unlikely | As Likely as Not | Somewhat Likely | Extremely Likely

Including a neutral midpoint is optional. Researchers sometimes omit it to encourage participants to make a definitive choice rather than defaulting to the middle. However, for bipolar scales (e.g., Agree-Disagree), a neutral midpoint can be helpful for those who genuinely feel neutral about the topic, such as that in Figure 9.6.3.

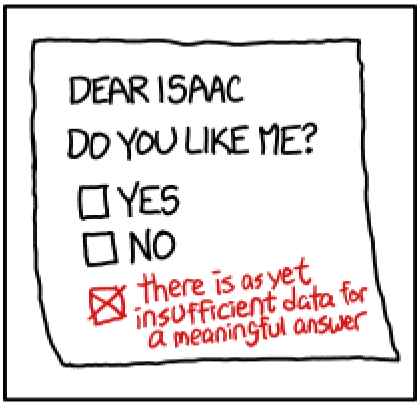


Figure 9.6.3. "Question" retrieved from http://imgs.xkcd.com/comics/question.png is used under a CC-BY-NC 2.5 licence

Formatting the Survey

Creating a well-organised survey goes beyond writing effective questions. The overall structure and presentation play a key role in ensuring respondents stay engaged and provide accurate answers.

Introduction:

Every survey should begin with a clear and engaging introduction that serves two main purposes. First, it should encourage participation. Unlike lab studies, where participants have already volunteered their time, survey respondents are often approached unexpectedly, such as through phone calls, emails, or mail. Researchers need to make a compelling case for why someone should take the time to respond. A strong introduction briefly explains the survey's purpose and significance, identifies the sponsor (e.g., a university, which often improves response rates), highlights the importance of each respondent's contribution, and mentions any incentives for participation.

Second, the introduction should establish informed consent. Respondents must understand the essential details before agreeing to participate. This includes the survey topics, estimated time commitment, confidentiality assurances, and the fact that they can withdraw at any time. While formal written consent forms are not always required in minimal-risk survey research, the introduction must still clearly communicate these points.

Instructions:

After the introduction, the survey should provide clear instructions for how to complete it. If the survey includes unusual response formats or rating scales, include simple examples to guide respondents. Clear instructions reduce confusion and improve data quality.

Order of Questions:

The order of the questions also matters. Respondents are typically most engaged and focused at the start of the survey, so begin with the most important questions, such as those that are directly tied to your research goals. Group similar items together; for example, keep questions using the same rating scale or focusing on a particular topic in one section. This approach makes it easier for respondents to follow and answer consistently.

Demographic Questions:

Save demographic questions (e.g., age, gender, income) for the end of the survey. These questions are generally less engaging but also quick and easy to answer, making them suitable for when respondents might be starting to feel fatigued.

A well-structured survey begins with a clear introduction, includes easy-to-follow instructions, presents important questions early, groups similar items together, saves demographics for the end, and finishes with a thank-you note.

References

Chang, L. and Krosnick, J. A. (2003). Measuring the frequency of regular behaviors: Comparing the "typical week" to the "past week". *Sociological Methodology, 33*(55-80). https://doi.org/10.1111/j.0081-1750.2003.t01-1-00127.x

Converse, J. M. (2009). *Survey research in the United States: Roots and emergence 1890-1960.* Routledge.

Krosnick, J. A., & Berent, M. K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science*, *37*(3), 941–964. https://doi.org/10.2307/2111580

Miller, J. M., & Krosnick, J. A. (1998). The impact of candidate name order on election outcomes. *The Public Opinion Quarterly*, *62*(3), 291–330. http://www.jstor.org/stable/2749662

Peterson, R. A. (2000). *Constructing effective questionnaires*. Sage Publications.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*(2), 93–105. https://doi.org/10.1037/0003-066X.54.2.93

Schwarz, N., & Strack, F. (1991). Context effects in attitude surveys: Applying cognitive theory to social research. *European Review of Social Psychology*, *2*(1), 31–50. https://doi.org/10.1080/14792779143000015

Strack, F., Martin, L. L. & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology, 18*(429-442). https://doi.org/10.1002/ejsp.2420180505

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* Jossey-Bass.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology , (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

9.7. QUASI-EXPERIMENTAL RESEARCH

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

The term quasi means "resembling", so quasi-experimental research is similar to true experimental research but lacks one key element. In a true experiment, random assignment ensures participants are evenly distributed across groups, or counterbalancing helps control order effects in within-subject designs. In quasi-experiments, one of these safeguards is missing. While the researcher manipulates an independent variable, either there is no control group, or participants are not randomly assigned to conditions (Cook & Campbell, 1979).

Quasi-experiments have an advantage over non-experimental research because the independent variable is manipulated before measuring the dependent variable. This eliminates the directionality problem, where it is unclear which variable affects the other. However, because random assignment or counterbalancing is not used, there is still a risk of confounding variables, which are other differences between groups that could explain the results. As a result, quasi-experiments fall somewhere between non-experimental research and true experiments in terms of internal validity.

Quasi-experiments are especially common in real-world field settings, where random assignment is impractical or impossible. They are frequently used to evaluate the effectiveness of treatments, such as a psychotherapy program or an educational intervention.

One-Group Posttest Only Design

In a one-group posttest only design, researchers apply a treatment or manipulation (independent variable) and then measure the dependent variable just once after the treatment. For example, a researcher might introduce an anti-drug education program to elementary school students and then immediately measure their attitudes toward illegal drugs after the program concludes.

This design is considered the weakest type of quasi-experimental design because it lacks a control or comparison group. Without a control group, there is no way to know what the students' attitudes would have been without the program. Any observed change could be due to other factors, such as natural development, outside influences, or participant expectations.

Despite these limitations, findings from one-group posttest only designs are often widely reported in the media and frequently misinterpreted by the public. For instance, an advertisement might claim that 80% of women noticed brighter skin after using a specific cleanser for a month. However, without a comparison group, it is impossible to know whether the improvement was actually caused by the cleanser or would have happened naturally over time.

One-Group Pretest-Posttest Design

In a one-group pretest-posttest design, researchers measure a dependent variable before and after introducing a treatment or intervention. For example, if a researcher wants to test the effectiveness of an anti-drug education program on elementary school students' attitudes toward illegal drugs, they might measure the students' attitudes one week, implement the program the next week, and then measure their attitudes again the following week.

This design is similar to a within-subjects experiment, where participants serve as their own control group. However, unlike a true experiment, there is no counterbalancing of conditions because participants cannot be exposed to the treatment before the control condition.

If students' attitudes improve after the anti-drug program, it seems logical to credit the program for the change. However, this conclusion is often uncertain because several threats to internal validity can provide alternative explanations for the observed change.

History is one such threat. Events occurring between the pretest and posttest might have influenced participants' attitudes. For example, a widely broadcast anti-drug message on TV or news of a celebrity's drug-related death could have affected the students' views independently of the program.

Maturation is another concern. Over time, participants naturally grow, learn, and develop. If the program lasted a year, improvements in reasoning skills or emotional maturity might explain the change, rather than the program itself.

The act of taking the pretest itself can also affect posttest results, a threat known as testing effects. For instance, completing a survey about drug attitudes might prompt participants to reflect on the topic, leading to changes in their attitudes before the program even begins.

Instrumentation can also undermine validity if the measuring tool changes over time. For example, participants might have been highly attentive during the pretest survey but less focused and engaged during the posttest, leading to inconsistent results.

Another issue is regression to the mean, a statistical phenomenon where individuals who score extremely high or low on one occasion are likely to score closer to the average on subsequent occasions. If participants were selected for the program based on unusually extreme attitudes toward drugs, their posttest scores would likely shift closer to average, regardless of the program's impact.

Closely related to regression to the mean is spontaneous remission, where medical or psychological problems improve naturally over time without treatment. For instance, people with depression often report improvement even without intervention. Research shows that participants in waitlist control groups for depression treatments tend to improve by 10–15% before receiving any therapy at all (Posternak & Miller, 2001).

Given these threats to internal validity, researchers must be cautious when interpreting results from one-group pretest-posttest designs. A common way to address these concerns is to add a control group, which is a group of participants who do not receive the treatment. Both groups would be subject to the same threats (e.g., history, maturation, testing effects), allowing researchers to more accurately measure the treatment's true effect.

However, adding a control group transforms the study into a two-group design, no longer qualifying as a one-group pretest-posttest design. While the one-group approach can offer useful preliminary insights, it is not sufficient for establishing strong causal conclusions.

Does Psychotherapy Work?

Early research on the effectiveness of psychotherapy often relied on pretest-posttest designs. In a landmark 1952 study, researcher Hans Eysenck reviewed data from 24 studies showing that about two-thirds of patients improved between the pretest and posttest. However, Eysenck compared these results with archival data from state hospitals and insurance company records, which showed that similar patients improved at roughly the same rate without receiving psychotherapy.

This comparison led Eysenck to suggest that the observed improvements might be due to spontaneous remission rather than psychotherapy itself. Importantly, Eysenck did not claim that psychotherapy was ineffective. Instead, he emphasised the need for carefully planned and well-executed experimental studies to evaluate psychotherapy's true effectiveness. His full article is available on the *Classics in the history of psychology* website.

Eysenck's call to action inspired further research. By the 1980s, hundreds of randomised controlled trials had been conducted, comparing participants who received psychotherapy with those who did not. In a 1980 meta-analysis, researchers Mary Lee Smith, Gene Glass, and Thomas Miller analysed these studies and concluded that psychotherapy is highly effective. Their results showed that approximately 80% of treatment participants improved more than the average participant in a control group.

Since then, research has shifted focus to understanding the specific conditions under which different types of psychotherapy are most effective.

Interrupted Time Series Design

The interrupted time series design is an extension of the pretest-posttest design, but with repeated measurements taken before and after a treatment over a period of time. A time series refers to a sequence of measurements recorded at consistent intervals, such as tracking weekly productivity in a factory for a year.

In an interrupted time series design, the regular time series is “interrupted” by a treatment or intervention. For example, in a classic study, researchers examined the effect of reducing factory work shifts from 10 hours to 8 hours (Cook & Campbell, 1979). After the shift reduction, productivity increased quickly and remained consistently high for several months. This pattern suggested that the shorter shifts were responsible for the improvement in productivity.

The key advantage of this design lies in its repeated measurements. Unlike a simple pretest-posttest design, which only measures the outcome once before and once after treatment, the interrupted time series includes multiple measurements both before and after the intervention. This allows researchers to detect trends, patterns, and variations over time, giving a clearer picture of the treatment’s effect.

Imagine a study measuring student absences in a research methods course. The treatment in this study is the instructor publicly taking attendance each day, making students aware that their presence or absence is being recorded.

In the top panel of Figure 9.7.1, the treatment is effective. Before attendance tracking begins, absences remain consistently high week after week. Once the instructor starts taking attendance, there is an immediate and lasting drop in absences. This pattern suggests that the treatment successfully encouraged students to attend class more regularly.

In the bottom panel of Figure 9.7.1, the treatment is ineffective. Despite starting public attendance tracking, the number of absences remains roughly the same before and after the treatment. This indicates that the intervention had little to no impact on attendance.

This example highlights a key advantage of the interrupted time-series design compared to a simpler pretest-posttest design. If researchers had only measured absences once before and once after the treatment, for example, at Week 7 and Week 8, it might have appeared that any change between these two weeks was caused by the intervention. However, the multiple measurements before and after reveal whether changes in attendance are part of a consistent trend or just random fluctuations.

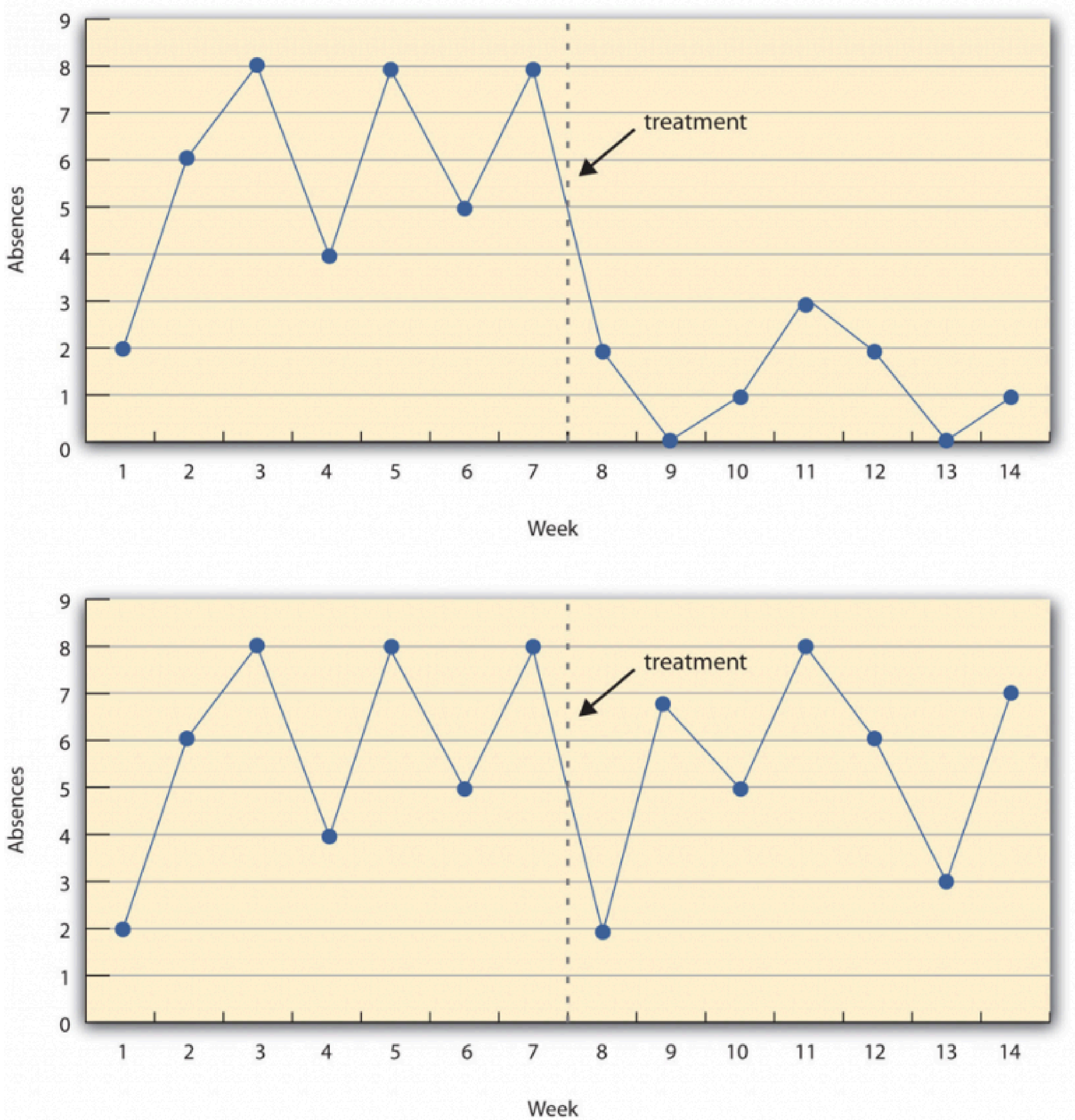


Figure 9.7.1. A hypothetical interrupted time-series design. The top panel shows data that suggest that the treatment caused a reduction in absences. The bottom panel shows data that suggest that it did not, by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

---

References

Cook, T. D., Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin.

Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: A meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders*, *66*(2), 139–146. https://doi.org/10.1016/S0165-0327(00)00304-9

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

9.8. FACTORIAL DESIGNS

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

In psychological research, it is common for studies to examine multiple levels of a single independent variable (e.g., placebo, new drug, old drug). However, it is equally common for experiments to include multiple independent variables within a single study. This approach allows researchers to explore more complex research questions. For example, Schnall and her colleagues (2008) studied whether physical disgust influences moral judgements by testing participants in either a clean or messy room and measuring their attention to bodily sensations ("private body consciousness"). Participants rated the moral acceptability of various behaviours and reported their current emotions. The study found that participants in the messy room felt more disgusted and made harsher moral judgements, but only if they had high private body consciousness. Hence, the researchers examined the effects of both disgust and private body consciousness in the same study.

The inclusion of multiple independent variables is further demonstrated by the following studies titles from professional journals:

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

Including multiple independent variables, just like using multiple levels of a single independent variable, allows researchers to address more sophisticated research questions. For instance, rather than conducting two separate studies, one on how disgust influences moral judgement and another on how private body consciousness affects moral judgement, Schnall and her colleagues combined these questions into a single study.

Beyond efficiency, this approach enables researchers to investigate whether the effect of one independent variable depends on the level of another. This phenomenon is known as an interaction between independent variables. For example, Schnall and colleagues found an interaction between disgust and private body consciousness: the effect of disgust on moral judgement was influenced by whether participants had high or low levels of private body consciousness. Interactions, such as this one, often produce some of the most intriguing results in psychological research.

Factorial Designs

A factorial design is a common approach in experiments involving multiple independent variables, often referred to as factors. In this design, every level of one independent variable is combined with every level of the other(s), creating all possible combinations of conditions. Each combination represents a unique condition in the experiment.

Example: 2 × 2 Factorial Design

Consider an experiment investigating the effect of cell phone use (yes vs. no) and time of day (day vs. night) on driving ability. This design is shown in the factorial table in Figure 9.8.1, where:

- Columns represent cell phone use.
- Rows represent the time of day.
- Cells represent the four combinations:
  - Using a cell phone during the day.
  - Not using a cell phone during the day.
  - Using a cell phone at night.
  - Not using a cell phone at night.

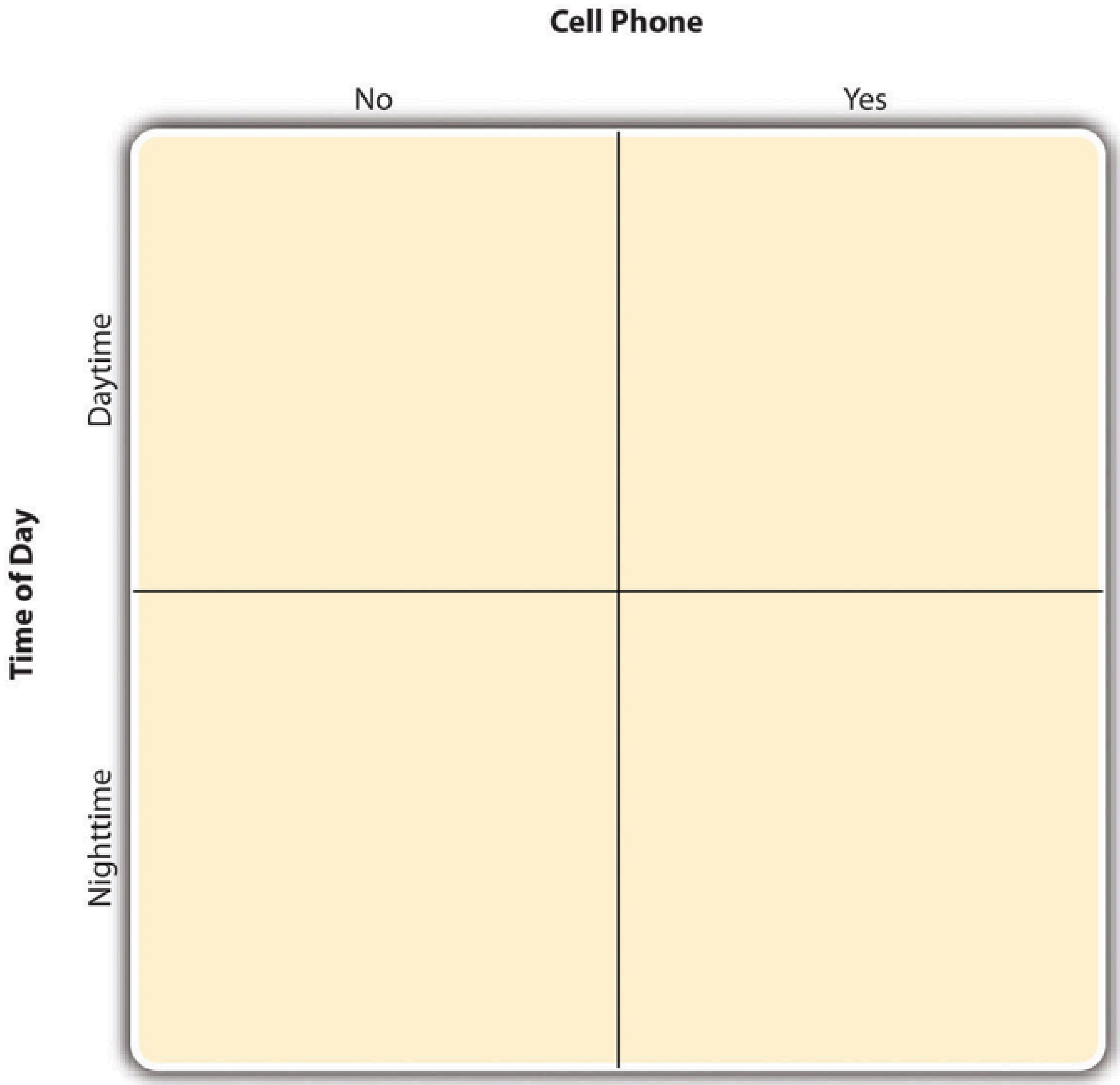


Figure 9.8.1. Factorial design table representing a 2 × 2 factorial design by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

This setup is called a 2 × 2 factorial design (read as "two-by-two") because it includes two independent variables, each with two levels.

Expanding Factorial Designs

If one variable includes a third level, for instance, differentiating between handheld cell phone use, hands-free cell phone use, and no cell phone use, the design becomes a 3 × 2 factorial design with six conditions. The total number of conditions is the product of the number of levels for each variable. For example:

- 2 × 2 design = 4 conditions.

- 3 × 2 design = 6 conditions.
- 4 × 5 design = 20 conditions.

The number of numbers in the factorial notation reflects the number of independent variables. For example:

- A 2 × 2 design has two independent variables, each with two levels.
- A 3 × 3 design also has two independent variables, but each has three levels.
- A 2 × 2 × 2 design has three independent variables, each with two levels.

Complex Designs and Practical Considerations

Factorial designs can theoretically include any number of independent variables with any number of levels. For example, a study might include:

- Type of psychotherapy (cognitive vs. behavioural),
- Length of therapy (2 weeks vs. 2 months),
- Therapist's sex (female vs. male).

This would be a 2 × 2 × 2 factorial design, resulting in eight unique conditions. Figure 9.8.2 represents such a design.

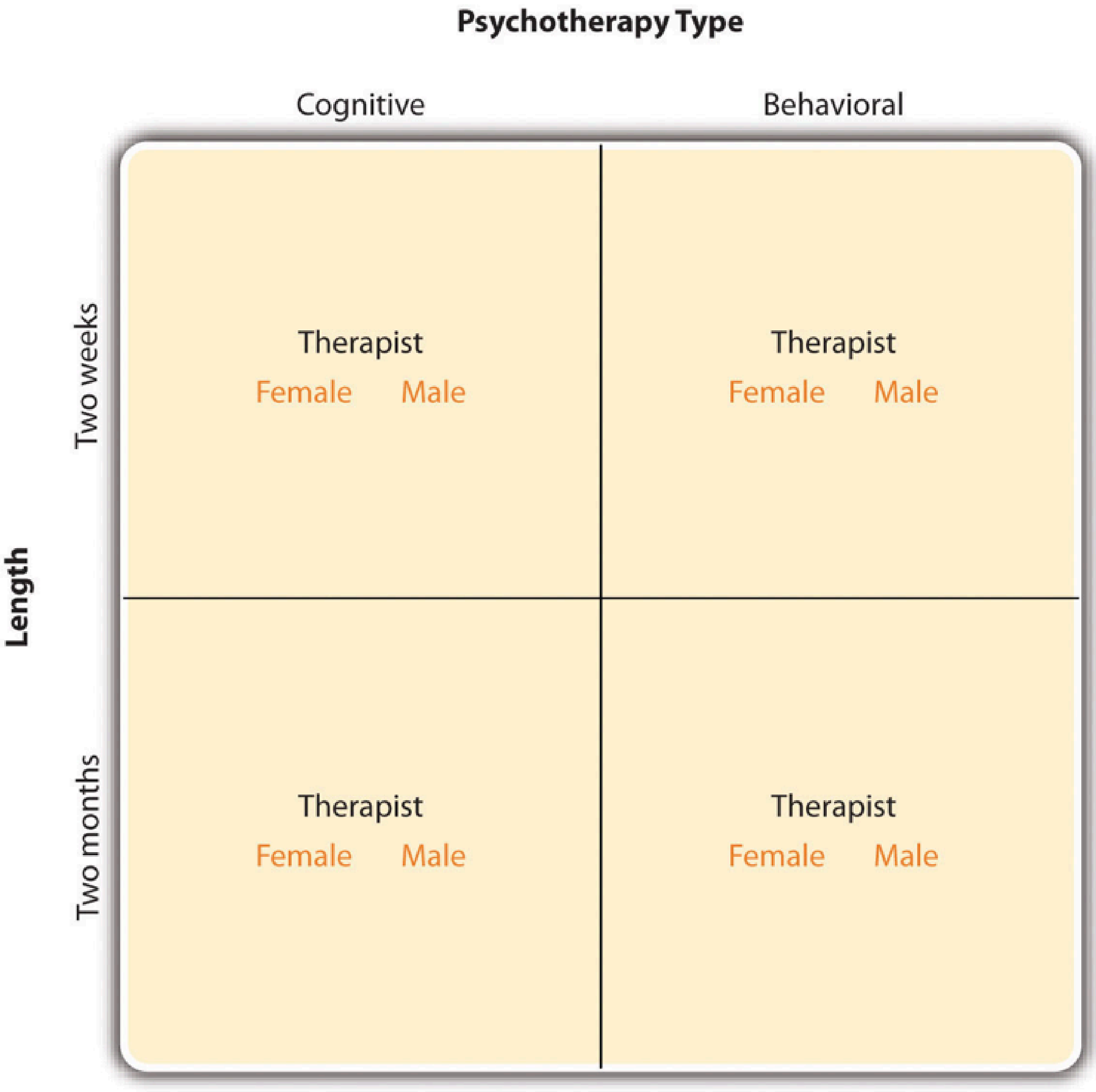


Figure 9.8.2. Factorial design table representing a 2 × 2 × 2 factorial design by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

However, adding more independent variables or levels can quickly make the design unmanageable:

1. **Increased Conditions:** Adding a fourth variable with three levels (e.g., therapist experience: low vs. medium vs. high) would create a 2 × 2 × 2 × 3 factorial design with 24 conditions.
2. **Participant Requirements:** More conditions require more participants to ensure adequate statistical power, making the design logistically and financially challenging.

For practical reasons, factorial designs typically involve no more than two or three independent variables, each with two or three levels. This simplifies the study while still allowing researchers to explore interactions between variables effectively. In this section, we focus on two-variable designs, though the principles discussed can be easily extended to more complex factorial designs.

Assigning Participants to Conditions

In a simple between-subjects design, each participant is tested in only one condition, whereas in a simple within-subjects design, each participant is tested in all conditions. In a factorial experiment, the decision to use a between-subjects or within-subjects approach must be made separately for each independent variable.

Between-Subjects Factorial Design

In a between-subjects factorial design, all independent variables are manipulated between subjects. For example, participants might be tested either while using a cell phone or not using a cell phone and either during the day or the night. Each participant would experience only one condition, ensuring they are tested in a single combination of the independent variables.

Advantages of this approach include its conceptual simplicity, the elimination of order or carryover effects, and the reduced time and effort required for participants. However, it often requires a larger number of participants to populate all conditions adequately.

Within-Subjects Factorial Design

In a within-subjects factorial design, all independent variables are manipulated within subjects. For example, participants might be tested both while using a cell phone and while not using a cell phone, and both during the day and the night. This means each participant would need to be tested in all four conditions.

The advantages of this design include greater efficiency for the researcher and better control over extraneous participant variables, such as individual differences. However, it may introduce order effects (where the sequence of conditions influences results) or carryover effects (where one condition affects performance in another). These issues can be mitigated through counterbalancing, where the order of conditions is varied systematically across participants.

Mixed Factorial Design

Since factorial designs involve multiple independent variables, it is also possible to use a mixed factorial design, where one independent variable is manipulated between subjects and another is manipulated within subjects.

For instance:

- Cell phone use could be a within-subjects factor, with participants tested both while using and not using a cell phone. Counterbalancing would ensure that the order of these two conditions varies across participants to prevent order effects.
- Time of day could be a between-subjects factor, with participants tested either during the day or at night. This approach might be chosen to minimise the burden on participants, as they would only need to attend a testing session once.

In this mixed design, each participant would experience two of the four possible conditions, combining the efficiency of within-subjects manipulation with the simplicity of between-subjects manipulation.

Regardless of whether a factorial design is between-subjects, within-subjects, or mixed, participants are typically assigned to conditions (or to the order of conditions) randomly. Random assignment ensures that differences between groups are distributed evenly, enhancing the validity of the study.

Non-Manipulated Independent Variables

In many factorial designs, one of the independent variables is non-manipulated. This means the researcher measures the variable but does not control or manipulate it. A study by Schnall and colleagues illustrates this distinction well. In their study, one independent variable, disgust, was manipulated by testing participants in either a clean or messy room. The other variable, private body consciousness, was a participant variable measured through a questionnaire.

Another example comes from a study by Halle Brown and colleagues, in which participants were asked to recall words they had previously seen (Brown et al., 1999). The manipulated independent variable was the type of word: some were health-related (e.g., tumour, coronary), while others were not health-related (e.g., election, geometry). The non-manipulated independent variable was participants' level of hypochondriasis (excessive concern about ordinary bodily symptoms), which was measured. The results showed that participants with high hypochondriasis were better at recalling health-related words but not better at recalling non-health-related words compared to those with low hypochondriasis.

Key Points About Non-Manipulated Independent Variables

1. **Typically Participant Variables:** Non-manipulated independent variables are usually participant variables, such as private body consciousness, hypochondriasis, self-esteem, or gender. These are inherently between-subjects factors, as participants cannot belong to more than one level of these variables (e.g., being both high and low in hypochondriasis).

2. **Experiments with Mixed Variables:** Studies with both manipulated and non-manipulated variables are still generally considered experiments, as long as at least one independent variable is manipulated. The inclusion of non-manipulated variables does not diminish the experimental nature of the study.
3. **Limits on Causal Conclusions:** Causal conclusions can only be drawn about manipulated independent variables. For instance, Schnall and colleagues could conclude that disgust influenced the harshness of moral judgements because it was manipulated, and participants were randomly assigned to the clean or messy room. However, they could not conclude that private body consciousness caused harsher moral judgements because it was a non-manipulated variable. A third variable, such as neuroticism, might underlie both private body consciousness and moral strictness.

Thus, it is crucial to distinguish between manipulated and non-manipulated variables in a study to correctly interpret which variables can support causal inferences and which cannot.

Non-Experimental Studies With Factorial Designs

While factorial experiments often include manipulated independent variables or a combination of manipulated and non-manipulated variables, they can also consist solely of non-manipulated independent variables. In these cases, the design is no longer experimental but rather non-experimental in nature.

For instance, imagine a study where a researcher measures the moods and self-esteem of participants, categorising them as having either a positive or negative mood and either high or low self-esteem. The researcher also measures participants' willingness to engage in unprotected sexual intercourse. This design can be conceptualised as a 2 × 2 factorial design, with mood (positive vs. negative) and self-esteem (high vs. low) as non-manipulated between-subjects factors. The dependent variable is participants' willingness to engage in unprotected sex.

Since neither independent variable is manipulated in this example, the study is classified as non-experimental. This contrasts with an experimental study, such as the one by MacDonald and Martineau (2002), where participants' moods were manipulated.

It is critical to exercise caution when interpreting results from non-experimental studies, particularly regarding causality. Non-experimental designs are vulnerable to issues like the directionality problem, where it is unclear whether one variable causes the other, and the third-variable problem, where an observed effect might be driven by another unmeasured variable. For example, if participants' moods appear to influence their willingness to engage in unprotected sex, this relationship might actually be explained by another factor correlated with mood, such as personality traits or situational influences.

Graphing the Results of Factorial Experiments

In factorial experiments with two independent variables, the results are typically graphed with one independent variable represented on the x-axis and the other variable distinguished by different-coloured bars or lines. The y-axis always represents the dependent variable.

Figure 9.8.3 illustrates this process using two hypothetical factorial experiments.

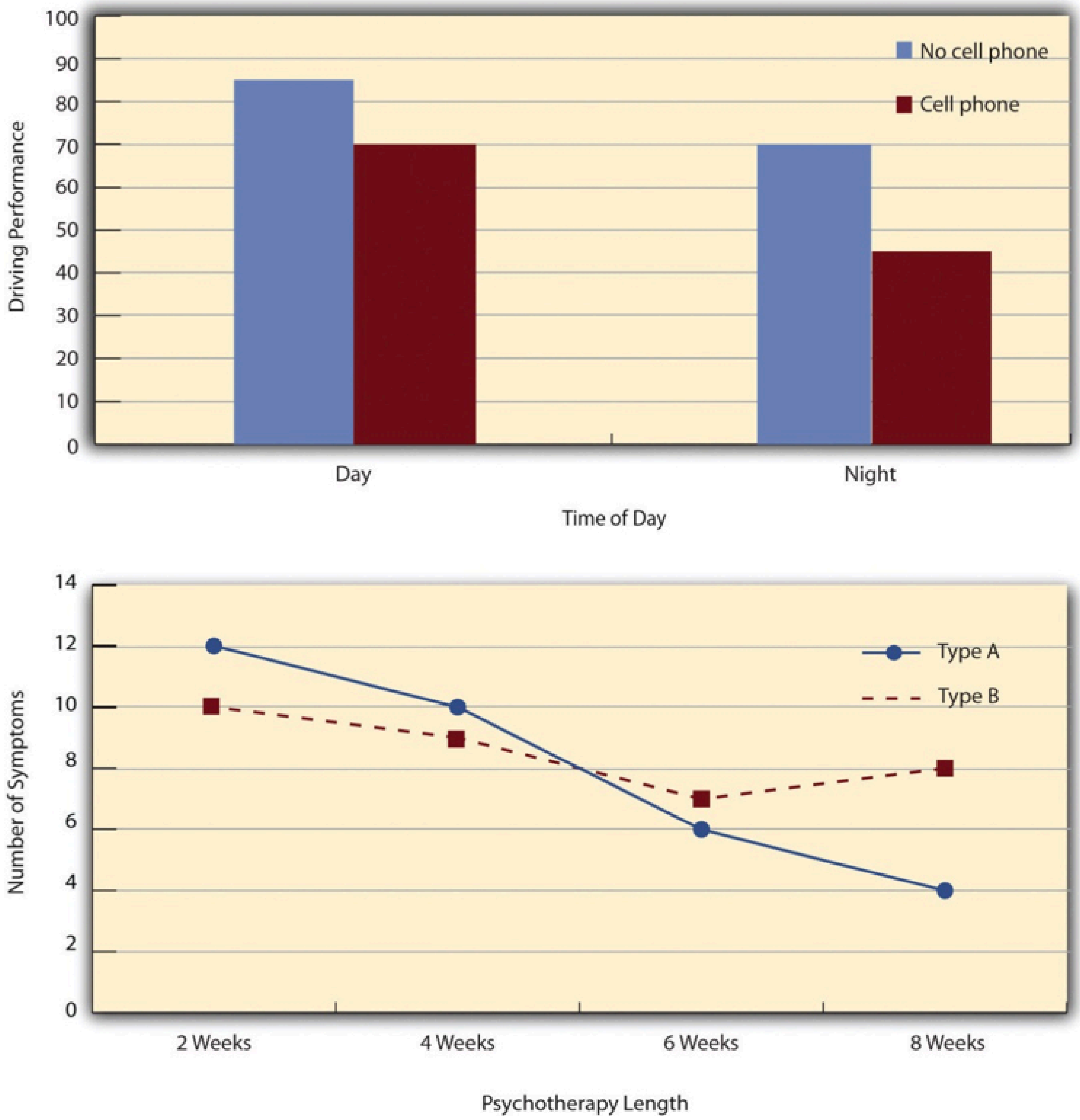


Figure 9.8.3. Two ways to plot the results of a factorial experiment with two independent variables by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

- **Top Panel:** This shows the results of a 2 × 2 factorial design, where time of day (day vs. night) is

plotted along the x-axis, and cell phone use (no vs. yes) is depicted using bars of different colours. Alternatively, cell phone use could be shown on the x-axis, with time of day represented by the bar colours. The decision depends on which configuration makes the data easier to interpret.
- **Bottom Panel:** This represents the results of a 4 × 2 factorial design, where one variable, psychotherapy length, is quantitative and displayed on the x-axis. The other variable, psychotherapy type, is indicated by differently formatted lines. In this case, a line graph is used instead of a bar graph because the x-axis variable is quantitative with a limited number of distinct levels. Line graphs are also ideal for representing measurements over time, referred to as time series data.

Main Effects

In factorial designs, researchers are interested in three types of results: main effects, interaction effects, and simple effects. A main effect refers to the impact of one independent variable on the dependent variable, averaged across the levels of the other independent variable. Each independent variable in the study has its own main effect to consider.

The top panel of Figure 9.8.3 illustrates a main effect of cell phone use because, on average, driving performance was better when participants were not using cell phones compared to when they were. This is evident as the blue bars (no cell phone use) are higher, on average, than the red bars (cell phone use). Similarly, there is a main effect of time of day, as driving performance was better during the day than at night, regardless of cell phone use.

Main effects are independent of each other, meaning the presence or absence of a main effect for one independent variable does not indicate whether there is a main effect for another. For example, the bottom panel of Figure 9.8.3 shows a clear main effect of psychotherapy length: the longer the psychotherapy, the better the outcome, regardless of the other variable in the study. This independence allows researchers to evaluate each variable's impact separately, providing a clearer understanding of the factors influencing the dependent variable.

Interactions

An interaction effect occurs when the effect of one independent variable depends on the level of another independent variable. While this may sound complex, the concept is quite intuitive. For example, imagine your friend invites you to a movie with another friend. Your response might be, "Well, it depends on which movie you're seeing and who else is coming." You are excited about the summer blockbuster but uninterested in the cheesy romantic comedy, which reflects a main effect of movie type. However, if your willingness to see the romantic comedy depends on whether your friend is bringing along someone you are interested in, there is an interaction. Similarly, drug interactions illustrate this concept. Viagra can be helpful for older men, and nitrates can effectively treat chest pain. However, taking both drugs together can be lethal, highlighting a critical interaction between the two medications.

Interactions are also prevalent in psychological research. For instance, the effect of receiving psychotherapy is stronger among people who are highly motivated to change compared to those who are not. This is an interaction because the effectiveness of psychotherapy depends on motivation levels. Similarly, Schnall and colleagues found an interaction in their study on moral judgements. The effect of a clean versus messy room on participants' moral judgements depended on their level of private body consciousness. Participants high in private body consciousness made harsher judgements in a messy room, whereas participants low in private body consciousness were unaffected by the room's condition.

In many studies, interactions are the primary focus of the research. For example, Brown and her colleagues hypothesised that individuals with hypochondriasis would recall negative health-related words more accurately than those without hypochondriasis but would recall non-health-related words at the same level as others. This hypothesis pointed to an interaction, and the results supported it.

Types of Interactions

The way one independent variable's effect depends on the level of another independent variable can manifest in several ways. Two common types of interactions are spreading interactions and crossover interactions.

Spreading Interactions

In a spreading interaction, the effect of one independent variable is present at one level of the other independent variable but is weaker or absent at the other level. The top two panels of Figure 9.8.4 illustrate this:

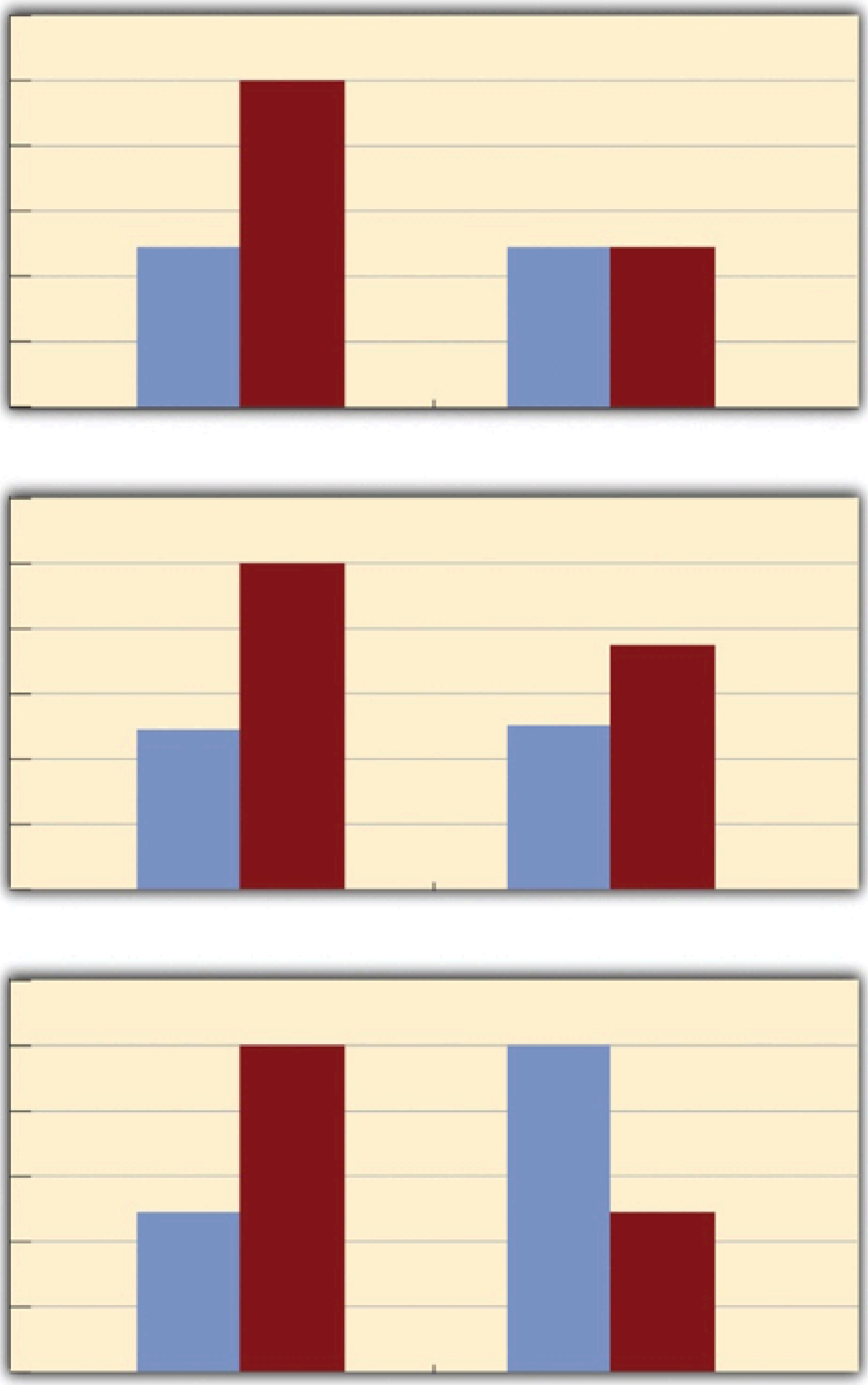


Figure 9.8.4. Bar graphs showing three types of interactions. In the top panel, one independent variable has an effect at one level of the second independent variable but not at the other. In the middle panel, one independent variable has a stronger effect at one level of the second independent variable than at the other. In the bottom panel, one independent variable has the opposite effect at one level of the second independent variable than at the other, by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

1. **Effect at One Level Only:** In the top panel, independent variable "B" has an effect at level 1 of independent variable "A" (indicated by the difference in the height of the blue and red bars on the left side of the graph). However, it has no effect at level 2 of independent variable "A" (the blue and

red bars are the same height on the right side). This pattern is similar to Schnall and colleagues' study, where disgust affected participants with high private body consciousness but had no effect on those with low private body consciousness.

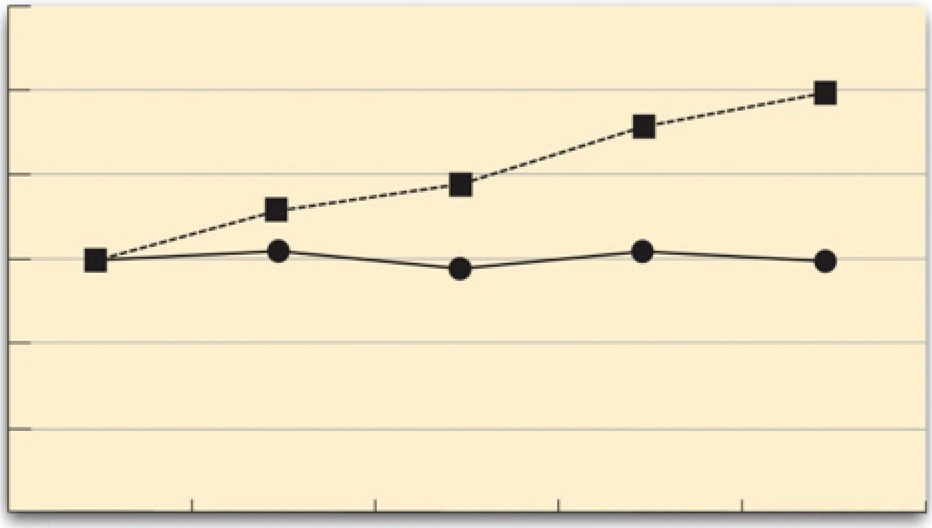
2. **Stronger Effect at One Level:** In the middle panel, independent variable "B" has a stronger effect at level 1 of independent variable "A" than at level 2. This is indicated by a larger difference in the height of the blue and red bars on the left side compared to a smaller difference on the right side. An example of this is a hypothetical driving study where using a cell phone had a strong effect at night but a weaker effect during the day.

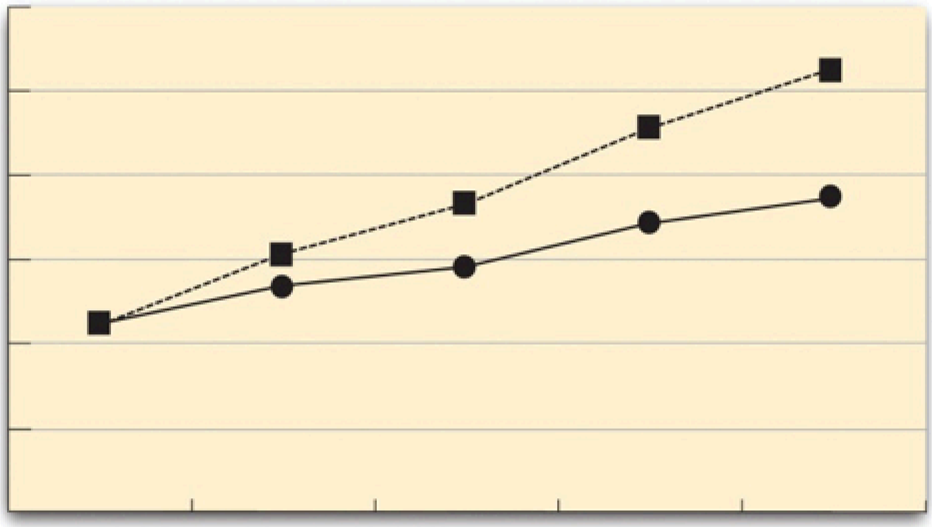
Crossover Interactions

A crossover interaction occurs when one independent variable affects both levels of the other independent variable, but the effects are in opposite directions. The bottom panel of Figure 9.8.4 demonstrates this: independent variable "B" produces opposite effects at levels 1 and 2 of independent variable "A."

A real-world example of a crossover interaction comes from Kathy Gilliland's study on the effect of caffeine on the verbal test scores of introverts and extraverts (Gilliland, 1980). Without caffeine, introverts perform better than extraverts. However, after consuming 4 mg of caffeine per kilogram of body weight, extraverts outperform introverts. This shows how the effect of caffeine depends on personality type.

Figure 9.8.5 provides examples of spreading and crossover interactions in line graphs where one independent variable is quantitative:





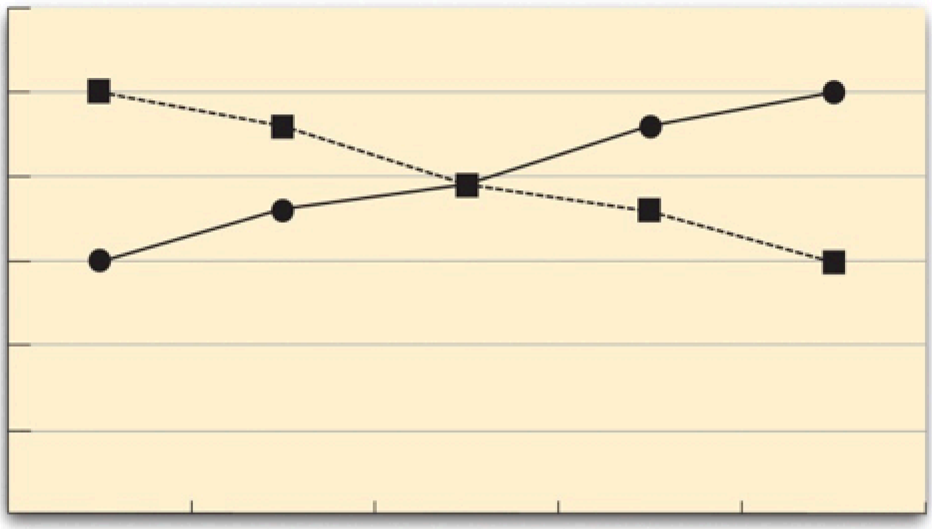


Figure 9.8.5. Line graphs showing different types of interactions. In the top panel, one independent variable has an effect at one level of the second independent variable but not at the other. In the middle panel, one independent variable has a stronger effect at one level of the second independent variable than at the other. In the bottom panel, one independent variable has the opposite effect at one level of the second independent variable than at the other, by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

- **Top Two Figures:** These illustrate two types of spreading interactions. In each, the effect of one independent variable diminishes or disappears at certain levels of the other independent variable.
- **Bottom Figure:** This depicts a crossover interaction, where the two lines literally "cross over," reflecting opposite effects of one independent variable at different levels of the other.

Simple Effects

When researchers find an interaction, it often reveals that the main effects may not tell the full story. Consider a crossover interaction where introverts outperform extraverts on a verbal memory test when they have not consumed caffeine, but extraverts outperform introverts when they have consumed caffeine. If researchers examine the main effect of caffeine consumption by averaging across introversion and extraversion, they might find no significant effect of caffeine overall, as the positive effect of caffeine for extraverts cancels out the negative effect for introverts. Similarly, examining the main effect of personality (introversion vs. extraversion) by averaging across caffeine conditions might show no overall effect, as the benefits of extraversion with caffeine balance out the drawbacks without it. However, the interaction suggests that the story is more complex because the effect of caffeine depends on personality.

To address this complexity, researchers use simple effects analyses, which break down the interaction to determine the effect of each independent variable at each level of the other independent variable. In this example, instead of averaging across personality, researchers would examine the effect of caffeine on introverts and then on extraverts. Similarly, they would analyse the effect of personality in the no-caffeine condition (where introverts performed better) and in the caffeine condition (where extraverts performed better). For a 2 × 2 design, this process results in two main effects and four simple effects analyses.

Examples of Simple Effects Analyses

Schnall and colleagues found a main effect of disgust on moral judgements, where participants in a messy room made harsher moral judgements overall. However, an interaction revealed that the effect of disgust depended on participants' level of private body consciousness. Simple effects analyses showed that for individuals high in private body consciousness, disgust significantly influenced moral judgements. Conversely, for those low in private body consciousness, there was no effect of disgust on moral judgements. These analyses provided a clearer understanding of the interaction and revealed that the main effect of disgust was only accurate for part of the sample.

Similarly, Brown and colleagues studied the interaction between word type (health-related vs. non-health-related) and hypochondriasis (high vs. low) on word recall. Simple effects analyses examined the effect of hypochondriasis on health-related words and non-health-related words separately. Results showed that participants high in hypochondriasis recalled more health-related words than those low in hypochondriasis, but there was no difference in recall for non-health-related words. This analysis highlighted the specific nature of the interaction.

When to Use Simple Effects Analyses

Simple effects analyses are only necessary when an interaction is present. If no interaction exists, main effects analyses provide a complete and accurate picture. Unlike main effects analyses, which average across levels of the other independent variable, simple effects analyses examine the effect of one independent variable at each level of the other.

For example:

- A 2 × 2 design with four conditions involves 2 main effects and 4 simple effects.
- A 2 × 3 design with six conditions involves 2 main effects and 5 simple effects.
- A 3 × 3 design with nine conditions involves 2 main effects and 6 simple effects.

The number of main effects is determined by the number of independent variables, while the number of simple effects depends on the levels of the independent variables. Each simple effect examines one independent variable's impact at a specific level of the other variable(s), providing detailed insights into how interactions shape the results.

References

Brown, H. D., Kosslyn, S. M., Delamater, B., Fama, J., & Barsky, A. J. (1999). Perceptual and memory biases for health-related information in hypochondriacal individuals. *Journal of Psychosomatic Research*, *47*(1), 67–78. https://doi.org/10.1016/S0022-3999(99)00011-2

Gilliland, K. (1980). The interactive effect of introversion-extraversion with caffeine induced arousal on verbal performance. *Journal of Research in Personality*, *14*(4), 482–492. https://doi.org/10.1016/0092-6566(80)90006-9

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

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, *34*(8), 1096-1109. https://doi.org/10.1177/0146167208317771

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

9.9. SUMMARY

By Marc Chao

Summary

Key experimental principles in psychology include manipulating independent variables, controlling extraneous variables, and using random assignment to ensure reliability and rule out alternative explanations. Experimental designs, such as single-factor two-level and multi-level setups, often include control conditions like placebos or wait-list groups to isolate treatment effects. The placebo effect demonstrates how expectations alone can influence outcomes, as shown in studies like placebo knee surgeries for osteoarthritis. Between-subjects experiments expose participants to only one condition, requiring random assignment to balance groups, while within-subjects designs test all conditions on the same participants, controlling variability but introducing order effects that can be mitigated through counterbalancing. Researchers must weigh these designs' benefits and limitations, selecting the most suitable approach based on the research question and practical considerations.

Validity is a cornerstone of psychological research, encompassing internal validity (establishing cause-and-effect relationships), external validity (generalisability), construct validity (accurately measuring intended concepts), and statistical validity (appropriate use of data analysis). While experiments excel in internal validity due to controlled conditions, external validity may be limited unless field studies are conducted. Construct validity relies on the alignment between experimental manipulations and research questions, while statistical validity requires proper sample sizes and analysis methods. Correlational research complements experiments by exploring statistical relationships between variables without manipulation, enabling description and prediction when experimental designs are impractical or unethical. However, its inability to establish causation due to directionality and third-variable problems underscores the importance of integrating experimental methods to reinforce theoretical insights.

Qualitative research offers a contrasting approach, focusing on understanding human experiences through rich, non-numerical data collection methods like interviews and thematic

analysis. This approach generates deep insights, often identifying patterns and hypotheses for future quantitative exploration. For instance, Lindqvist et al. used unstructured interviews with families of teenage suicide victims to uncover the enduring emotional impact of unanswered questions. Qualitative research captures lived experiences, while mixed-methods designs bridge the depth of qualitative insights with the statistical rigour of quantitative analysis, providing a comprehensive understanding of human behaviour.

Survey research exemplifies flexibility in psychological studies, relying on self-reports and often large random samples to describe variables or assess relationships. With roots in early 20th-century social surveys, innovations like the Likert scale have expanded its applications to topics like mental health prevalence and public attitudes. Surveys can be experimental or non-experimental, requiring careful design to minimise biases and errors. Features like randomisation and clear question ordering enhance reliability, while qualitative and quantitative analyses enable a balance between depth and breadth. Mixed-methods approaches often combine these strengths to capture complex phenomena effectively.

Quasi-experimental research, which lacks random assignment or counterbalancing, addresses directionality problems by manipulating independent variables but remains vulnerable to confounding variables. Common designs like one-group posttest-only and pretest-posttest setups offer preliminary insights but face threats to internal validity, such as history and maturation effects. Interrupted time series designs strengthen causal interpretations by using repeated measurements before and after interventions, distinguishing treatment effects from random fluctuations. These designs are valuable in real-world settings where true experiments are impractical but require cautious interpretation to avoid overstating causal claims.

Factorial designs allow researchers to explore complex interactions between multiple independent variables, combining their levels into unique conditions. For example, Schnall et al. demonstrated how disgust (manipulated by room cleanliness) and private body consciousness interacted to influence moral judgements. These designs, such as 2 × 2 or more complex setups, investigate main effects and interactions, with simple effects analyses offering nuanced insights into how variables interact. While factorial designs enhance understanding of multifactorial influences, they require careful planning to manage complexity and control confounding variables.

CHAPTER 10: DESCRIPTIVE STATISTICS

Descriptive statistics form the foundation of data analysis, providing essential tools for summarising, organising, and interpreting data. In psychology, these techniques help researchers identify patterns, trends, and relationships, laying the groundwork for deeper statistical exploration. This chapter delves into the key components of descriptive statistics, including the distribution of variables, measures of central tendency, and measures of variability, each offering unique insights into the nature of the data.

Understanding a variable's distribution is crucial for identifying how values are spread across categories or levels. Frequency tables and histograms present these distributions clearly, helping researchers spot common values, ranges, and outliers. Measures of central tendency, including the mean, median, and mode, offer different ways to pinpoint the dataset's "typical" value, while measures of variability, such as range, standard deviation, and variance, reveal the extent to which data points cluster around or deviate from the centre.

This chapter also explores how to describe and interpret statistical relationships between variables, whether comparing groups or examining correlations between quantitative measures. Finally, the importance of effectively presenting descriptive statistics is emphasised, with guidelines for using text, figures, and tables in alignment with APA standards. Through clear organisation and systematic analysis, descriptive statistics provide a comprehensive snapshot of the data, setting the stage for meaningful research conclusions.

Learning Objectives

By the end of this chapter, you should be able to:

- **Define descriptive statistics:** Explain what descriptive statistics are and their role in summarising, organising, and displaying data in psychological research.
- **Explain variable distributions:** Define a variable's distribution and interpret how its values are spread across participants using examples.
- **Recognise distribution shapes:** Identify common distribution shapes, including unimodal, bimodal, symmetrical, positively skewed, and negatively skewed patterns.
- **Understand measures of central tendency:** Define and calculate the mean, median, and mode, and explain when each measure is most appropriate based on data

distribution and outliers.

- **Understand measures of variability:** Explain the range, variance, and standard deviation, and interpret what these measures reveal about data spread.
- **Interpret percentile ranks and *z*-scores:** Define percentile ranks and *z*-scores, calculate them, and explain their significance in comparing individual scores within a dataset.
- **Describe statistical relationships between variables:** Differentiate between relationships based on group differences and correlations between quantitative variables, and explain the importance of effect sizes such as Cohen's *d* and correlation coefficients like Pearson's *r*.
- **Create and interpret data visualisations:** Use bar graphs, line graphs, and scatterplots to represent and compare data, and understand the role of error bars and regression lines in these visualisations.
- **Present descriptive statistics effectively:** Follow APA guidelines for presenting descriptive statistics clearly and consistently in writing, tables, and figures.
- **Prepare and organise data for analysis:** Explain the importance of securing, organising, and formatting raw data for analysis, and understand best practices for ensuring data accuracy and clarity.
- **Differentiate between planned and exploratory analyses:** Explain the difference between planned (hypothesis-driven) and exploratory (data-driven) analyses and the importance of transparency when reporting these approaches.

10.1. THE DISTRIBUTION OF A VARIABLE

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

A variable's distribution shows how its values are spread out across different categories or levels. In simple terms, it tells us how often each value occurs in the data.

For example, imagine a survey of 100 university students measuring the variable "number of siblings". The distribution might look like this:

- 10 students have no siblings
- 30 students have one sibling
- 40 students have two siblings

Similarly, for a variable like "sex," the distribution might show:

- 44 students identify as male
- 56 students identify as female

A variable's distribution gives us a clear picture of how its values are shared among participants, making it easier to understand and interpret the data.

Frequency Tables

A frequency table is a simple and clear way to show how data is distributed across different values of a variable. It organises data into two columns: one for the possible values of the variable and the other for how often each value appears in the dataset.

For example, consider Table 10.1.1, which displays scores from the Rosenberg Self-Esteem Scale for 40 college students. The first column lists the self-esteem scores, and the second column shows how many students received each score. For instance, three students scored 24, five scored 23, and so on. From this table, we can easily observe:

- The range of scores (from 15 to 24)

- The most common score (22) and the least common score (17)
- Any unusual or extreme values

Table 10.1.1. Frequency table showing a hypothetical distribution of scores on the Rosenberg Self-Esteem Scale

Self-Esteem Score	Frequency
24	3
23	5
22	10
21	8
20	5
19	3
18	3
17	0
16	2
15	1

Key Features of Frequency Tables

Order of Values: The values in the first column are typically arranged from highest to lowest and only include scores that actually appear in the dataset. For example, although Rosenberg scores can range from 0 to 30, the table above only includes scores from 15 to 24 because that is where the data lies.

Grouped Frequency Tables: When a dataset includes many different scores or a wide range of values, it is more practical to group scores into ranges. In a grouped frequency table, the first column shows score ranges, while the second column lists how many scores fall into each range. For example, Table 10.1.2 displays grouped reaction times for 20 participants.

Table 10.1.2. A grouped frequency table showing a hypothetical distribution of reaction times

Reaction Time (ms)	Frequency
241–260	1
221–240	2
201–220	2
181–200	9
161–180	4
141–160	2

In grouped tables:

- The ranges must have equal widths (e.g., all intervals in Table 10.1.2 span 20 milliseconds).
- There are typically between 5 and 15 ranges for clarity.

Categorical Variables: Frequency tables can also be used for categorical variables, where the values in the first column are category labels instead of numerical scores (e.g., gender, favourite colour). In these cases, the order of the categories is usually based on frequency, starting with the most common category at the top.

Histograms

A histogram is a visual representation of how data is distributed across different values of a variable. It shows the same information as a frequency table, but in a format that is often quicker and easier to understand.

In a histogram:

- The x-axis represents the values of the variable (e.g., self-esteem scores).
- The y-axis represents how often each value occurs (frequency).
- Each value or range of values is represented by a vertical bar. The height of the bar corresponds to the number of individuals with that score or within that range.

When the variable is quantitative (e.g., self-esteem scores, reaction times), the bars are placed side by side with no gaps between them. This indicates a continuous range of values. However, if the variable is categorical (e.g., gender, favourite colour), there are usually small gaps between the bars to show that the values represent separate categories.

For example, if we create a histogram to represent the Rosenberg Self-Esteem Scores from Table 10.1.1, the x-axis would show the range of self-esteem scores (e.g., 15 to 24), and the y-axis would show how many students scored at each level (Figure 10.1.1). Each score would have a corresponding bar, and the bar for a score like 17 would be absent or have a height of zero if no students had that score.

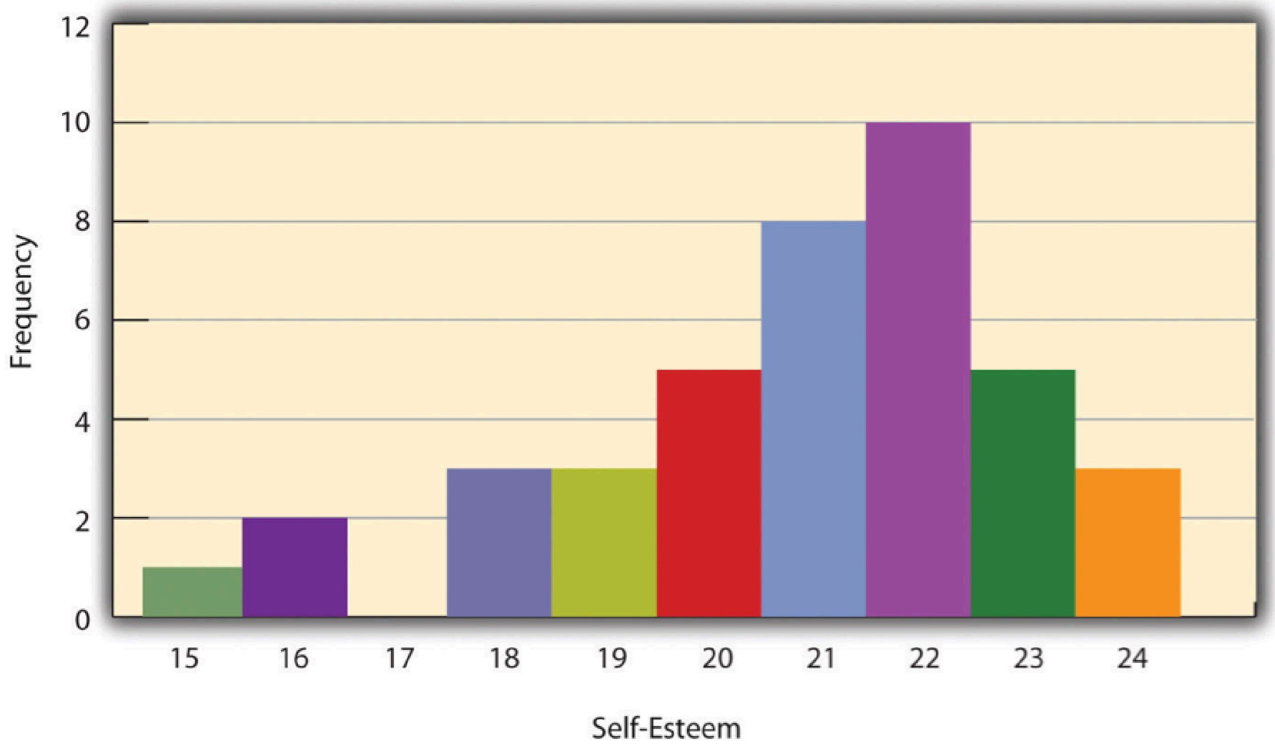


Figure 10.1.1. Histogram showing the distribution of self-esteem scores presented in Table 10.1.1 by R. S. Jhangiani et al. in [Research Methods in Psychology 4e] is used under a [CC BY-NC-SA licence]

Distribution Shapes

When data from a quantitative variable is shown in a histogram, the overall arrangement of the bars creates a shape. This shape helps us understand how the data is spread across different values and reveals patterns or trends.

Peaks in Distributions

A common distribution shape has a single peak near the middle, with the bars gradually tapering off on both sides. This type of distribution is called unimodal, meaning it has one clear high point. For example, a histogram showing self-esteem scores might have most values clustering around a central point, forming one peak.

Sometimes, a distribution has two distinct peaks, which is called bimodal. This can happen if the data naturally clusters into two groups. For example, the histogram in Figure 10.1.2 shows hypothetical scores on the Beck Depression Inventory with two peaks: one for individuals with low depression scores and another for those with high depression scores. Although distributions can have more than two peaks (multimodal), such patterns are rare in psychological research.

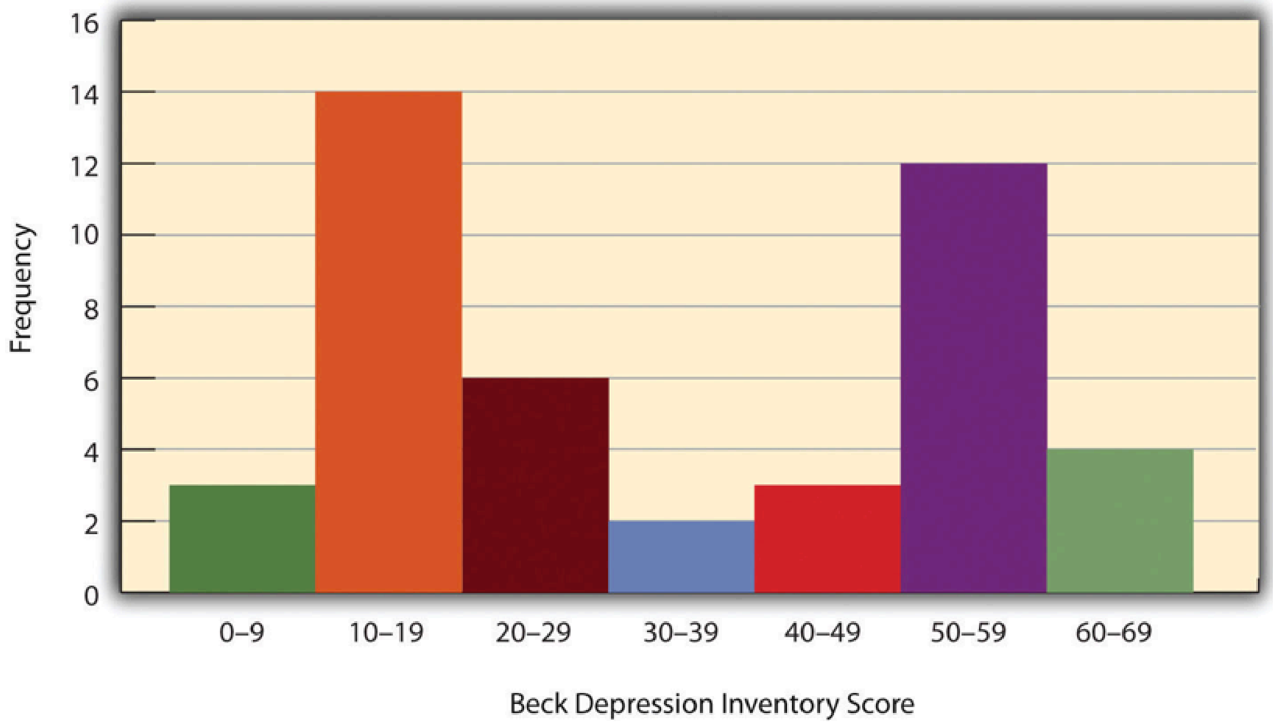


Figure 10.1.2. Histogram showing a hypothetical bimodal distribution of scores on the Beck Depression Inventory by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Symmetrical vs. Skewed Distributions

In Figure 10.1.3, we see another important feature of a distribution's shape is whether it is symmetrical or skewed:

- A symmetrical distribution has two sides that are mirror images of each other, with the peak in the centre.
- A negatively skewed distribution has a peak shifted toward the higher end of the range, with a long tail stretching toward the lower values.
- A positively skewed distribution has a peak shifted toward the lower end of the range, with a long tail stretching toward the higher values.

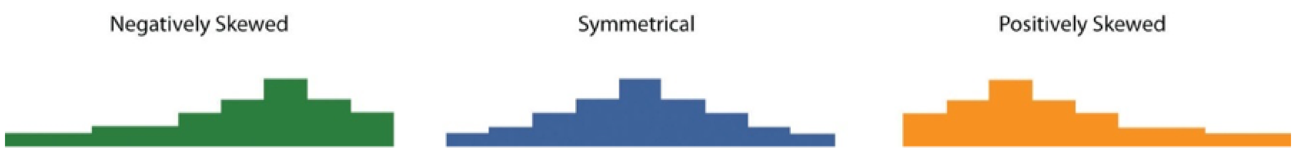


Figure 10.1.3. Histograms showing negatively skewed, symmetrical, and positively skewed distributions by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Skewness often indicates something meaningful about the data. For example, a negatively skewed distribution might suggest a test was too easy, while a positively skewed distribution could indicate a test was too difficult.

Outliers

An outlier is a data point that is much higher or lower than the rest of the scores in the dataset. Outliers can occur for several reasons:

- They may represent genuine extreme values. For example, one person in a sample might score extremely high on a depression inventory while the rest score low.
- They can also result from errors, such as incorrect data entry, misinterpretation by participants, or equipment malfunctions.

Outliers are important because they can distort statistical results and misrepresent trends in the data. Later in this chapter, we will discuss how to identify, interpret, and handle outliers effectively.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

10.2. MEASURES OF CENTRAL TENDENCY

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Central tendency refers to the middle point or centre of a set of data, which is a value around which the scores in a dataset tend to cluster. In simpler terms, it represents the "average" or typical value in a distribution. For example, if we look back at the self-esteem scores in Figure 9.1, we can see that most scores are clustered around the values 20 to 22. There are three main ways to measure central tendency: the mean, the median, and the mode.

The Mean (*M*)

The mean is the most common measure of central tendency and is often referred to as the average. It is calculated by adding up all the scores and then dividing the total by the number of scores. Mathematically, it is expressed as:

$$M = \frac{\Sigma X}{N}$$

- Σ (sigma): Represents the summation.
- X: Represents each individual score. Hence, ΣX means to sum across the values of the variable X.
- N: Represents the total number of scores.

The mean is widely used because it is easy to calculate, easy to understand, and has statistical properties that make it valuable for advanced analysis. However, the mean can be misleading in datasets with extreme scores (outliers) because those extreme values can significantly affect the result.

The Median

The median is the middle score in a dataset when the scores are arranged in order from lowest to highest. It splits the data so that half the scores are below it and half are above it.

To find the median:

1. Arrange the scores from lowest to highest.
2. Identify the middle value.

Example: Consider the dataset: 8, 4, 12, 14, 3, 2, 3.

Step 1: Arrange the scores in order → 2, 3, 3, 4, 8, 12, 14.
Step 2: Identify the middle score → The median is 4 because it is the middle value with three scores below and three scores above.

When there is an even number of scores, the median is the mean average of the two middle scores. For example, if we add a score of 15 to the dataset, we now have: 2, 3, 3, 4, 8, 12, 14, 15. The middle two scores are 4 and 8, so the median would be:

- (4 + 8) / 2 = 6

The median is less affected by outliers than the mean, making it a better measure of central tendency for skewed datasets.

The Mode

The mode is the most frequently occurring score in a dataset. It identifies the value that appears the most often.

For example, in the self-esteem dataset shown in Table 9.1 and Figure 9.1, the mode is 22, as more students had this score than any other.

The mode is unique because:

- It can be used for both quantitative and categorical data.
- A dataset can have more than one mode (bimodal or multimodal).

Comparing the Mean, Median, and Mode

In a symmetrical, unimodal distribution (a dataset with one clear peak), the mean, median, and mode are typically very close to each other, sitting around the peak.

In a bimodal distribution (a dataset with two peaks), the mean and median fall between the two peaks, while the mode aligns with the peaks.

In a skewed distribution:

- A positively skewed distribution (tail extends to the right): The mean is pulled toward the higher scores, making it larger than the median.
- A negatively skewed distribution (tail extends to the left): The mean is pulled toward the lower scores, making it smaller than the median.

Example of Skewed Data:

Consider these reaction times (in milliseconds): 200, 250, 280, 250.

- The mean is 245 ms.

If we add one outlier of 5,000 ms (an unusually long delay), the mean jumps to 1,445 ms, even though most scores are much lower. In such cases, the median (which remains unaffected by the outlier) is a better representation of central tendency.

Which Measure Should You Use?

Each measure of central tendency provides a different perspective on the data:

- **Mean:** Best for normally distributed data without outliers.
- **Median:** Best for skewed data or when outliers are present.
- **Mode:** Useful for identifying the most frequent value and can be applied to categorical data.

You do not have to rely on just one measure. Often, using multiple measures together gives a clearer and more accurate picture of the data's central point.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

10.3. MEASURES OF VARIABILITY

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Variability tells us how spread out or clustered the scores are around the central point of a dataset. Even if two datasets have the same mean, median, and mode, they can still differ greatly in how much the scores vary. For example, Figure 10.3.1 shows two groups of students who both score an average of 10 on a test. In one group, most students score close to 10, while in the other group, scores are scattered widely between 2 and 18. These differences in variability reveal important information about the data.

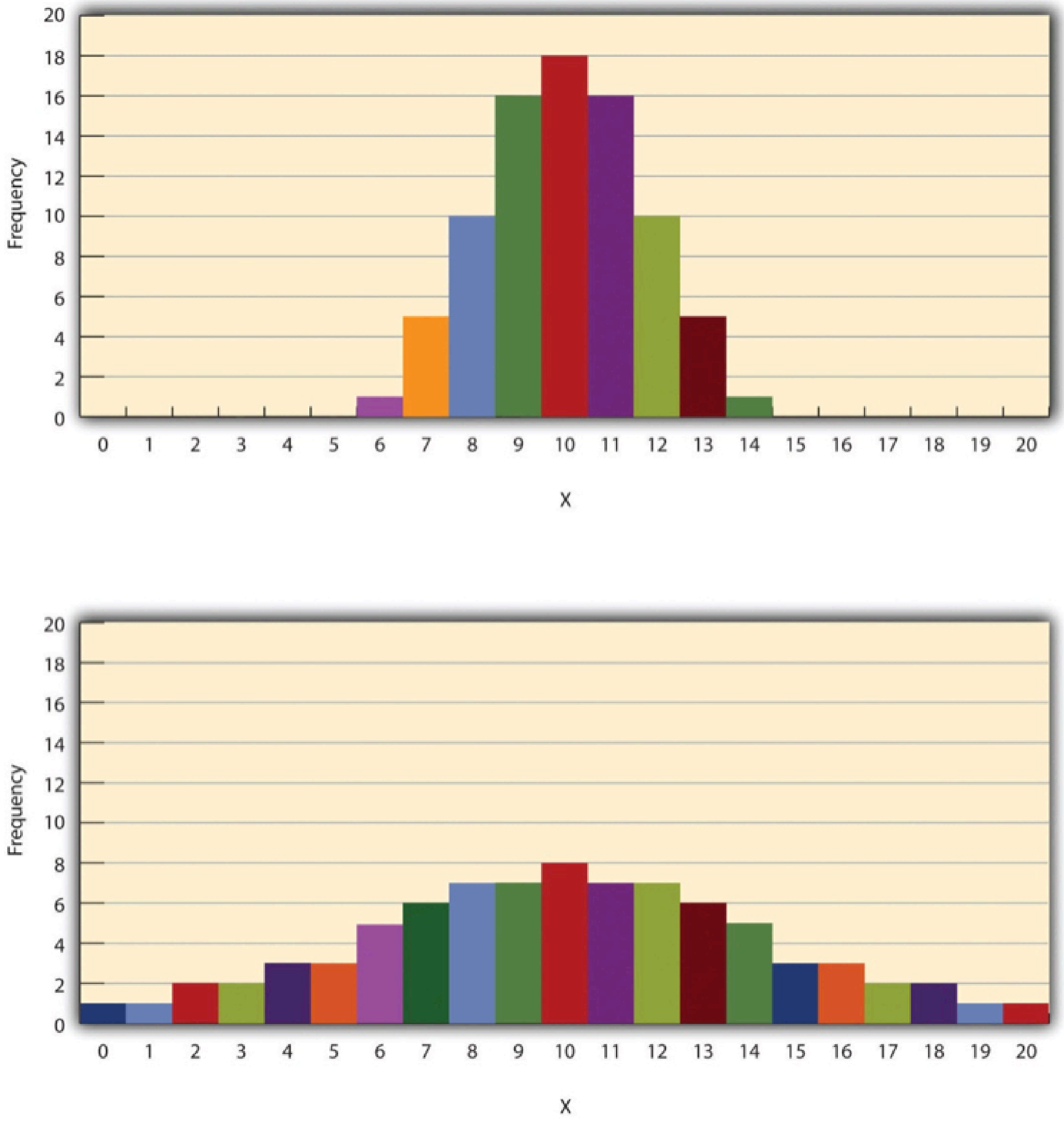


Figure 10.3.1. Histograms showing hypothetical distributions with the same mean, median, and mode but with low variability (top) and high variability (bottom) by R. S. Jhangiani et al. in [Research Methods in Psychology 4e]() is used under a [CC BY-NC-SA licence]()

The Range

The range is the simplest measure of variability. It is calculated by subtracting the lowest score from the highest score:

- Range = Highest Score – Lowest Score

For example, if the highest self-esteem score is 24 and the lowest is 15, the range would be:

- 24 – 15 = 9

While the range is easy to calculate and understand, it has a significant limitation: it is heavily influenced by outliers (extremely high or low scores). For instance, if most students score between 90 and 100 on an exam but one student scores 20, the range increases dramatically, making the data seem more variable than it actually is.

The Standard Deviation

The standard deviation is the most commonly used measure of variability because it gives a more precise picture of how scores are spread around the mean. It measures the average distance between each score and the mean.

For example:

- In a dataset with low variability, most scores will be close to the mean, resulting in a small standard deviation.
- In a dataset with high variability, scores will be spread out far from the mean, leading to a large standard deviation.

In Figure 10.3.1:

- The top distribution has a standard deviation of 1.69, showing that most scores are close to the mean.
- The bottom distribution has a standard deviation of 4.30, indicating that the scores are more widely spread out.

Calculating the standard deviation involves several steps:

1. Find the difference between each score and the mean.
2. Square each difference.
3. Calculate the mean of these squared differences (this is called the variance).
4. Take the square root of the variance.

Mathematically, it looks like this:

$$SD = \sqrt{\frac{\sum(X-M)^2}{N}}$$

- ***SD*:** Standard deviation
- **Σ (Sigma):** Sum of all values
- ***X*:** Individual score
- ***M*:** Mean of the scores
- ***N*:** Total number of scores

This formula might look complicated, but it simply ensures that every score's distance from the mean is considered, and squaring the differences eliminates negative values.

Variance

In the process of calculating the standard deviation, there is an intermediate step called the variance (symbolized as SD^2). Variance is the mean of the squared differences from the mean. While it is less intuitive than the standard deviation, variance is important for more advanced statistical techniques, especially in inferential statistics.

For practical purposes in descriptive statistics:

Standard deviation is the preferred measure because it is in the same units as the original data (e.g., if data is in seconds, the standard deviation is also in seconds).
Variance, on the other hand, is expressed in squared units, which makes it harder to interpret directly.

Percentile Ranks and z-Scores

When analysing data, it is often helpful to know where an individual score falls within a larger dataset. Two common ways to describe a score's position are percentile ranks and z-scores.

Percentile Ranks

A percentile rank shows what percentage of scores in a dataset are lower than a specific score.

For example:

- In a dataset of 40 self-esteem scores (as shown in Table 10.1.1), suppose five students scored 23.
- By counting how many students scored lower than 23, we find that 32 students (80%) had lower scores.
- Therefore, a score of 23 corresponds to the 80th percentile, meaning those students scored higher than 80% of their peers.

Percentile ranks are especially common in standardised testing. For instance:

- If your percentile rank on a verbal ability test is 40, it means you scored higher than 40% of the people who took the test.

Percentiles provide a quick and easy way to compare individual performance against a larger group.

z-Scores

While percentile ranks tell us about a score's relative position, z-scores offer a precise measurement of how far a score is from the mean, expressed in standard deviation units.

The formula for calculating a z-score is:

$$z = \frac{(X - M)}{SD}$$

- ***X*:** The individual score
- ***M*:** The mean of the dataset
- ***SD*:** The standard deviation

For example:

In an IQ score dataset where:

- Mean (*M*) = 100
- Standard deviation (*SD*) = 15

An individual with an IQ score of 110 would calculate their z-score as:

$$z = \frac{(110 - 100)}{15} = +0.67$$

This means the score is 0.67 standard deviations above the mean.

Similarly, an IQ score of 85 would calculate as:

$$z = \frac{(85 - 100)}{15} = -1.00$$

This means the score is 1 standard deviation below the mean.

z-scores are valuable because they:

- **Show a Score's Position in a Dataset:** They clearly indicate how far a score is from the average in standardised units.
- **Identify Outliers:** Scores with z-scores below −3.00 or above +3.00 are often considered outliers because they fall more than three standard deviations from the mean.
- **Enable Comparisons Across Different Datasets:** Because z-scores standardise data, they allow comparisons between different datasets with different means and standard deviations.
- **Support Advanced Statistical Calculations:** z-scores are often used as building blocks for other statistical analyses, which we will explore later.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

10.4. DESCRIBING STATISTICAL RELATIONSHIPS

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

In psychological research, most questions focus on exploring the relationships between variables. These relationships help researchers understand patterns, make predictions, and test theories about behaviour, thoughts, and emotions.

Statistical relationships generally fall into two main categories:

- **Differences between groups or conditions:** These relationships compare two or more groups or experimental conditions to see if they differ in some measurable way. For example, a study might compare self-esteem scores between individuals who practice daily meditation and those who do not.
- **Relationships between quantitative variables:** These relationships examine how two numerical variables change together. For instance, a researcher might investigate whether higher levels of stress are linked to lower academic performance.

In this section, we will take a closer look at how to describe, interpret, and present these statistical relationships in a clear and meaningful way. Understanding these relationships allows researchers to draw conclusions and contribute valuable insights to the field of psychology.

Differences Between Groups or Conditions

When researchers compare groups or conditions in psychological studies, they often describe the results using two key statistical measures: the mean and the standard deviation for each group. These measures help summarise the average performance of each group and show how much individual scores vary around that average.

For example, Thomas Ollendick and his colleagues (2009) conducted a study to compare two treatments for children with simple phobias, such as an intense fear of dogs. The children were randomly assigned to one of three groups:

1. **Exposure Group:** Children directly confronted their fears with the help of a trained therapist.

2. **Education Group:** Children learned about phobias and coping strategies.
3. **Wait-List Control Group:** Children received no treatment during the study but were promised treatment afterwards.

To measure the severity of the children's phobias, a clinician (who did not know which treatment each child received) rated their fear on a scale from 1 to 8. The results showed:

- **Exposure Group:** Mean = 3.47, Standard Deviation = 1.77
- **Education Group:** Mean = 4.83, Standard Deviation = 1.52
- **Control Group:** Mean = 5.56, Standard Deviation = 1.21

These results indicate that both treatments helped reduce fear, but the exposure treatment was more effective than the education treatment. Differences like these are often illustrated using bar graphs, like the one shown in Figure 10.4.1, where each bar represents the mean score for a group.

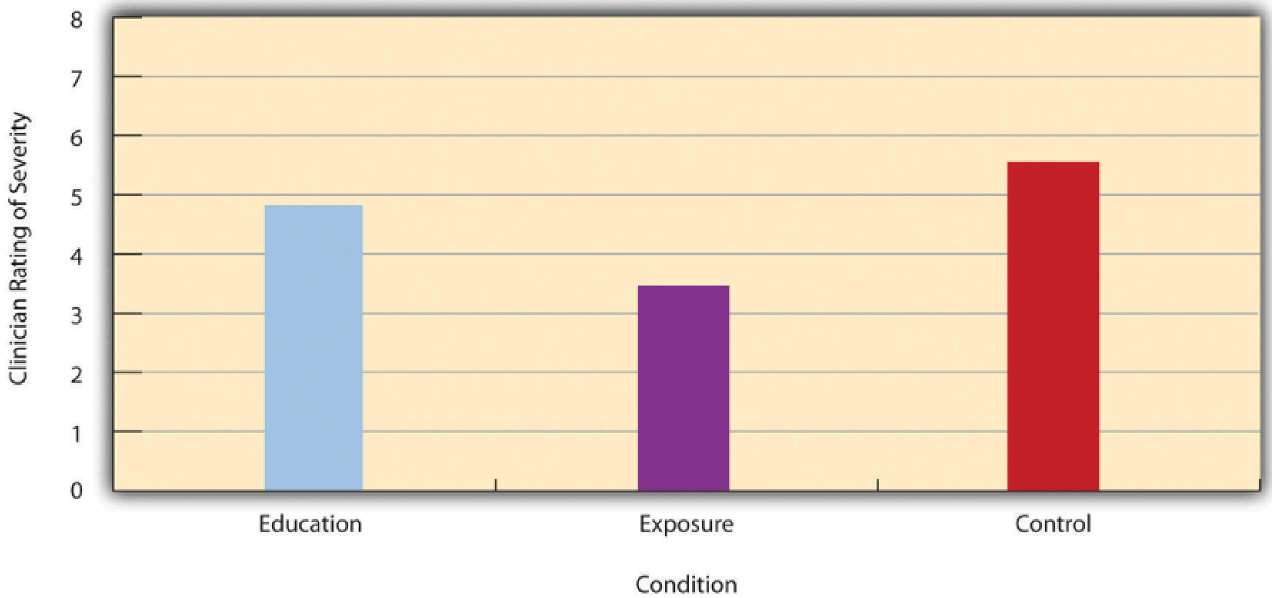


Figure 10.4.1. Bar graph showing mean clinician phobia ratings for children in two treatment conditions by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Effect Size: Cohen's *d*

In addition to reporting the mean and standard deviation, researchers often calculate the effect size to describe the strength of the difference between groups. The most common measure of effect size for group comparisons is Cohen's *d*, calculated using this formula:

$$d = \frac{M_1 - M_2}{SD}$$

- M_1 and M_2 represent the means of the two groups.
- *SD* represents the standard deviation (often an average of the two groups' standard deviations, known as the pooled standard deviation).

Cohen's *d* tells us how much the two group means differ in terms of standard deviation units. For example:

- A Cohen's *d* of 0.50 means the group means differ by half a standard deviation.
- A Cohen's *d* of 1.20 means the group means differ by 1.2 standard deviations.

To interpret Cohen's *d* values:

- 0.20 indicates a small effect size
- 0.50 indicates a medium effect size
- 0.80 or higher indicates a large effect size

In Ollendick's study, the difference between the exposure and education treatments had a large effect size ($d = 0.82$), highlighting a significant difference in their effectiveness.

Cohen's d is valuable because it provides a standardised measure of the difference between groups. Whether measuring self-esteem, reaction times, or blood pressure, a Cohen's *d* of 0.20 always indicates a small effect, and a *d* of 0.80 always indicates a large effect. This standardisation allows researchers to:

- Compare results across different studies.
- Combine data from multiple studies in meta-analyses.
- Communicate findings more clearly.

Caution About the Term 'Effect Size'

It is important to note that effect size does not automatically imply causation. For example:

- In an experiment where participants are randomly assigned to exercise or no-exercise groups, a Cohen's *d* of 0.35 might suggest that exercise caused a slight increase in happiness.
- In a cross-sectional study, where researchers simply compare people who exercise with those who do not, the same effect size would only suggest an association, not a cause-and-effect relationship.

While effect size is a powerful tool for understanding group differences, it must be interpreted carefully within the context of the study design.

Correlations Between Quantitative Variables

In psychology, many research questions focus on understanding relationships between quantitative variables. These relationships, called correlations, help researchers identify patterns and trends in data.

For example, researchers Kurt Carlson and Jacqueline Conard (2011) studied the link between the alphabetical order of people's last names and their response speed to consumer offers. In their study, MBA students were emailed about free basketball tickets available in limited supply. The results in Figure 10.4.2 showed that students with last names closer to the end of the alphabet tended to respond more quickly.

These relationships are often shown visually using line graphs or scatterplots.

- **Line Graphs:** These are used when the x-axis represents a variable with a small number of distinct values, like the four quartiles of last names in Carlson and Conard's study (as shown in Figure 10.4.2). Each point on the line graph shows the average response time for a group of students based on their alphabetical quartile.

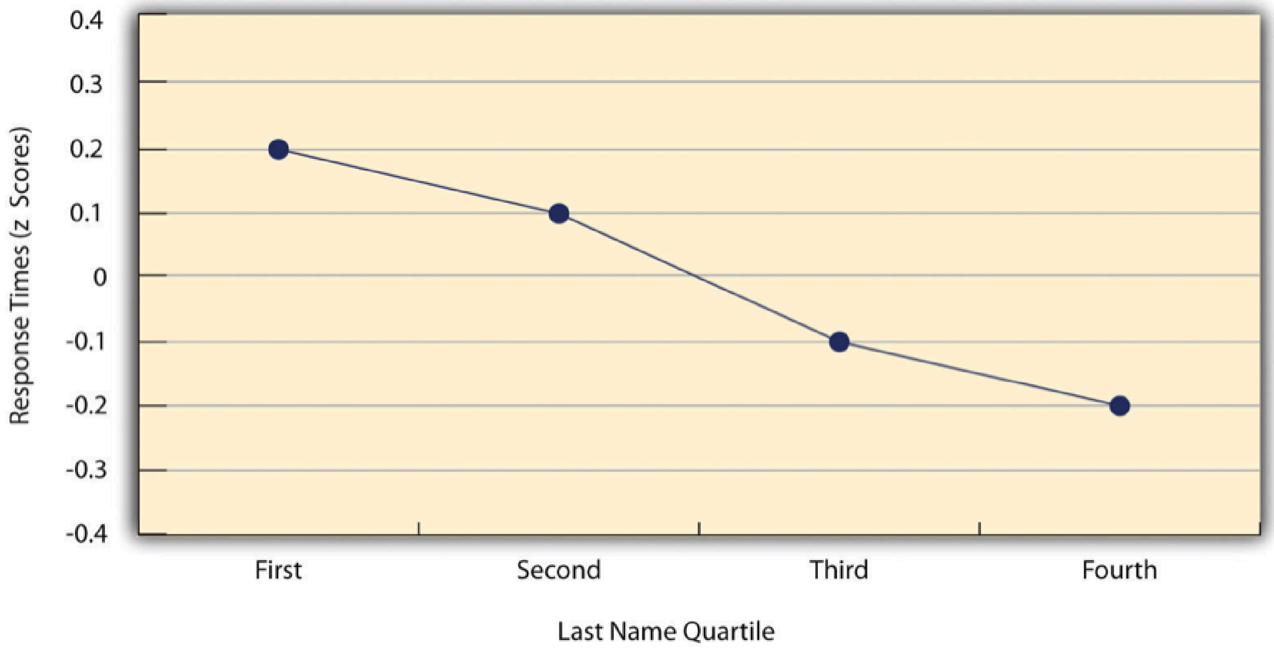


Figure 10.4.2. Line graph showing the relationship between the alphabetical position of people's last names and how quickly those people respond to offers of consumer goods by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

- **Scatterplots:** These are used when the x-axis represents a variable with many possible values, such as self-esteem scores. Each point represents an individual participant. For example, a scatterplot might show the relationship between students' self-esteem scores on two different occasions (as shown in

Figure 10.4.3).

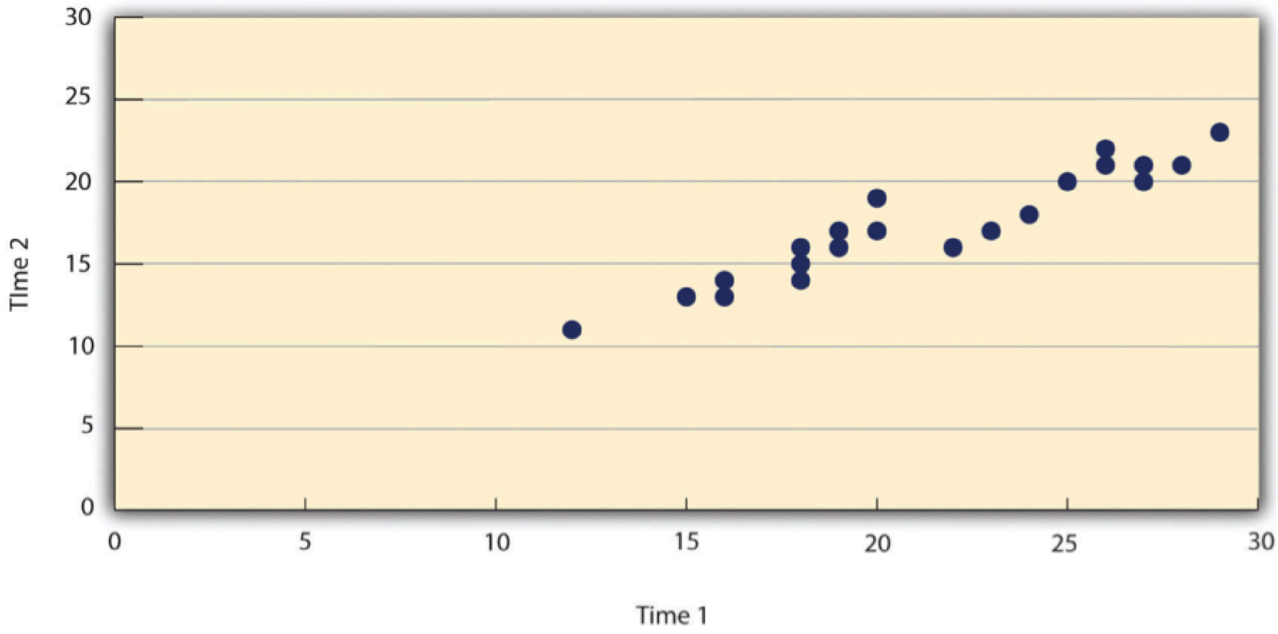


Figure 10.4.3. Statistical relationship between several university students' scores on the Rosenberg Self-Esteem Scale given on two occasions a week apart by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Positive vs. Negative Relationships

Correlations can show positive or negative relationships:

- In a positive relationship, higher scores on one variable are associated with higher scores on another. On a scatterplot, the points tend to trend upward from bottom-left to top-right.
- In a negative relationship, higher scores on one variable are associated with lower scores on another. The points trend downward from top-left to bottom-right.

Both of these relationships are considered linear relationships because they can be represented by a straight line on a scatterplot. Sometimes, however, relationships are nonlinear, meaning the data points follow a curved pattern instead.

For example, the relationship between sleep duration and depression levels might form an upside-down U-shaped curve. People who sleep around eight hours may have lower depression levels, while those who sleep too little or too much might have higher levels of depression. In such cases, a straight line cannot accurately represent the data (as shown in Figure 10.4.4).

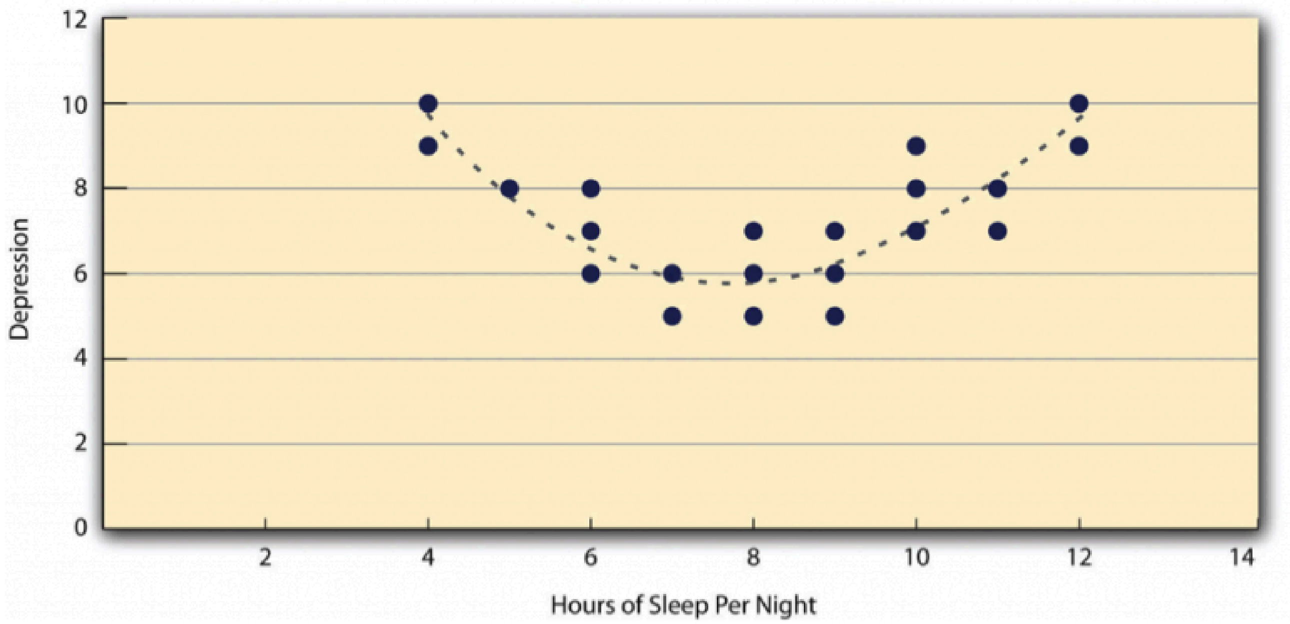


Figure 10.4.4. A hypothetical nonlinear relationship between how much sleep people get per night and how depressed they are by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Measuring the Strength of Correlations: Pearson's *r*

As discussed in Chapter 6, to quantify the strength and direction of a correlation, researchers use Pearson's *r*, a statistical measure that ranges from -1.00 to +1.00:

- +1.00 represents a perfect positive relationship.
- -1.00 represents a perfect negative relationship.
- 0.00 means no relationship exists between the two variables.

According to Cohen's guidelines, as illustrated in Figure 10.4.5:

- ±0.10 indicates a small correlation.
- ±0.30 indicates a medium correlation.
- ±0.50 indicates a large correlation.

It is important to remember that the sign (+ or -) indicates the direction of the relationship, not its strength. For example, +0.30 and -0.30 are equally strong, but one shows a positive trend and the other a negative trend.

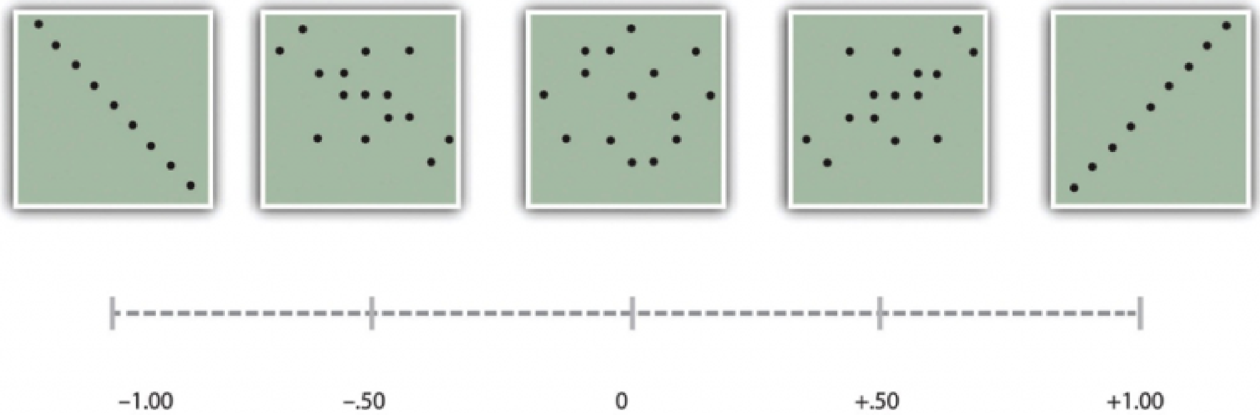


Figure 10.4.5. Pearson's r ranges from −1.00 (representing the strongest possible negative relationship), through 0 (representing no relationship), to +1.00 (representing the strongest possible positive relationship) by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

How Pearson's *r* is Calculated

While software usually handles these calculations, understanding the process helps clarify what Pearson's *r* represents. It is essentially the average of the cross-products of *z*-scores for two variables:

$$r = \frac{\sum(z_x z_y)}{N}$$

1. Convert each score into a *z*-score by subtracting the mean and dividing by the standard deviation for that variable.
2. Multiply each pair of *z*-scores (one from each variable).
3. Find the average of these cross-products.

This calculation results in a value between -1.00 and +1.00, summarising the strength and direction of the relationship.

When Pearson's *r* Can Be Misleading

- **Nonlinear Relationships:** Pearson's *r* assumes a linear relationship. If the data follows a curve (like the sleep-depression example), Pearson's *r* might inaccurately suggest no relationship. Always check your scatterplot to ensure the relationship is roughly linear before relying on Pearson's *r*.
- **Restriction of Range:** Pearson's *r* can also be misleading if the range of one or both variables is limited. For example in Figure 10.4.6:
  - A study might show a strong negative correlation between age and enjoyment of hip-hop music across all age groups.
  - However, if the sample includes only 18- to 24-year-olds, the correlation might appear weak or

non-existent because the variability in age is too limited.

  ◦ To avoid this issue, researchers should aim to collect data from a wide range of values for each variable. If restriction of range occurs, Pearson's *r* should be interpreted cautiously.

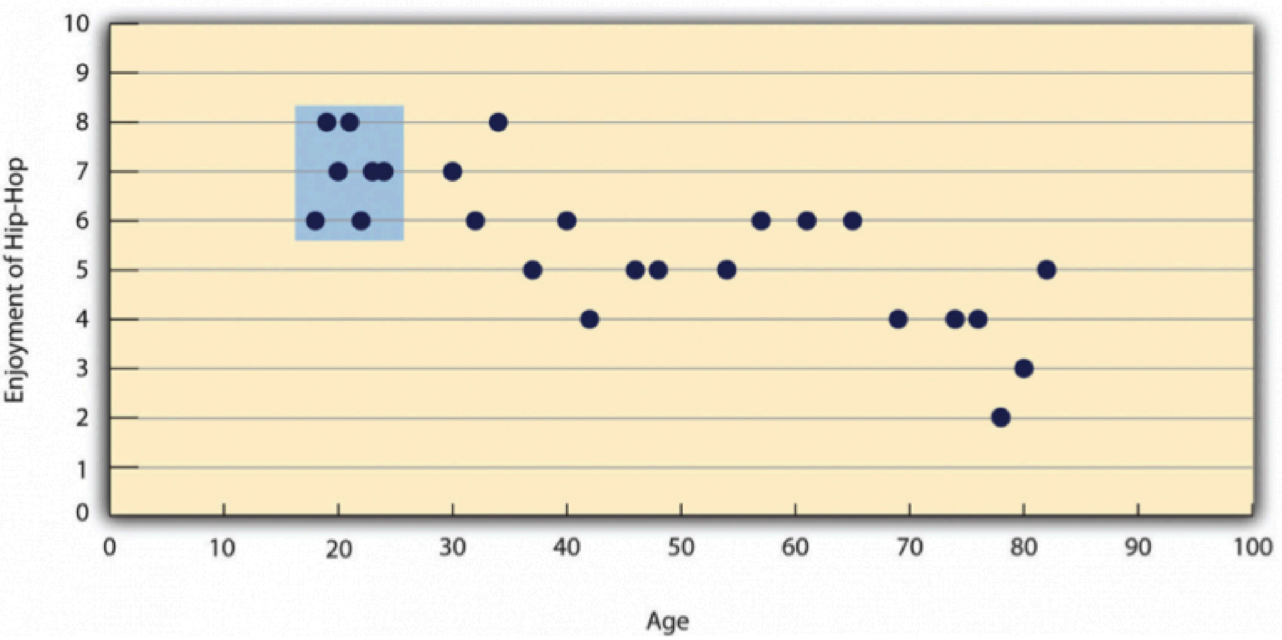


Figure 10.4.6. Hypothetical data showing how a strong overall correlation can appear to be weak when one variable has a restricted range. The overall correlation here is −.77, but the correlation for the 18- to 24-year-olds (in the blue box) is 0, by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

---

References

Carlson, K. A., & Conard, J. M. (2011). The last name effect: How last name influences acquisition timing. *The Journal of Consumer Research*, *38*(2), 300–307. https://doi.org/10.1086/658470

Ollendick, T. H., Öst, L.-G., Reuterskiöld, L., Costa, N., Cederlund, R., Sirbu, C., Davis, T. E. III, & Jarrett, M. A. (2009). One-session treatment of specific phobias in youth: A randomized clinical trial in the United States and Sweden. *Journal of Consulting and Clinical Psychology*, *77*(3), 504–516. https://doi.org/10.1037/a0015158

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

10.5. PRESENTING YOUR RESULTS

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

After analysing your data using descriptive statistics, the next step is to communicate your findings clearly and effectively. This section will guide you on how to present your results in writing, figures, and tables, following the American Psychological Association (APA) guidelines. These guidelines ensure consistency and clarity, whether you are preparing a written research report, a poster, or a slideshow presentation.

Presenting Descriptive Statistics in Writing

When presenting descriptive statistics in APA style, clarity and consistency are essential.

First, when writing numbers in text, words should be used for numbers less than 10, provided they do not represent precise statistical values. For numbers 10 and above, as well as all statistical results, numerals should always be used. Statistical results must also be presented as numerals (e.g., 2.00 instead of two or 2), and they should typically be rounded to two decimal places, unless specified otherwise. Results can either be included directly in the narrative or placed in parentheses, similar to how reference citations are handled.

When reporting a small number of results, embedding them directly into the text is often the most effective approach. For example, the mean age of the participants was 22.43 years, with a standard deviation of 2.34. Another example might read: Participants with low self-esteem in a negative mood expressed stronger intentions to have unprotected sex (M = 4.05, SD = 2.32) compared to those in a positive mood (M = 2.15, SD = 2.27). Similarly, the treatment group had a mean score of 23.40 (SD = 9.33), while the control group had a mean score of 20.87 (SD = 8.45). Additionally, the test-retest correlation was 0.96, or There was a moderate negative correlation between the alphabetical position of participants' last names and their response time (r = −0.27).

When results are integrated into the narrative, the terms mean and standard deviation should be written out in full. However, when they are included parenthetically, the symbols M and SD should be used instead. Maintaining consistency in style is particularly important when presenting comparable results. For example:

- ✔️ 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 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 reporting a large amount of data, figures, such as pie charts, bar graphs, and scatterplots, can often present information more clearly and efficiently than text alone (Figure 10.5.1) . In an APA-style research report, these graphical representations are referred to as figures.

When preparing figures, it is essential to follow some general principles to ensure clarity and accuracy. First, figures should add value to the presentation of results rather than simply repeat information already provided in text or tables. If a figure makes the data clearer or more efficient to understand, you might consider removing the corresponding text or table. Second, figures should be kept as simple as possible. While colour can be effective in posters, slideshow presentations, or textbooks, the APA Publication Manual recommends avoiding unnecessary use of colour in printed reports unless it adds essential clarity. Third, figures should be self-explanatory. Readers should be able to understand the main findings directly from the figure and its caption without having to refer to the main text.

In addition to these general principles, there are specific technical guidelines for creating figures according to APA style:

Graph Layout:

- Graphs, including scatterplots, bar graphs, and line graphs, should generally be slightly wider than they are tall for better readability.
- The independent variable should always be plotted on the x-axis, while the dependent variable goes on the y-axis.
- Values on the x-axis should increase from left to right, and values on the y-axis should increase from bottom to top.
- Both axes should ideally start at zero unless a different starting point is necessary for clarity.

Axis Labels and Legends:

- Axis labels should be clear, concise, and include units of measurement if these are not already specified in the caption.
- Axis labels should run parallel to their respective axes for clarity.
- Legends should appear within the figure and should be easy to interpret.
- The font style and size used in the figure should be consistent throughout, with a size no smaller than 8 points and no larger than 14 points.

Captions:

- Every figure caption should begin with the word "Figure", followed by its number in the order it appears in the text, ending with a period. The title should be italicised.
- The caption should include a brief description of the figure, followed by a period (e.g., "Reaction times of the control versus experimental group.").
- Any additional information necessary to interpret the figure, such as abbreviations, units of measurement (if not already specified on the axes), or units for error bars, should also be included after the description.

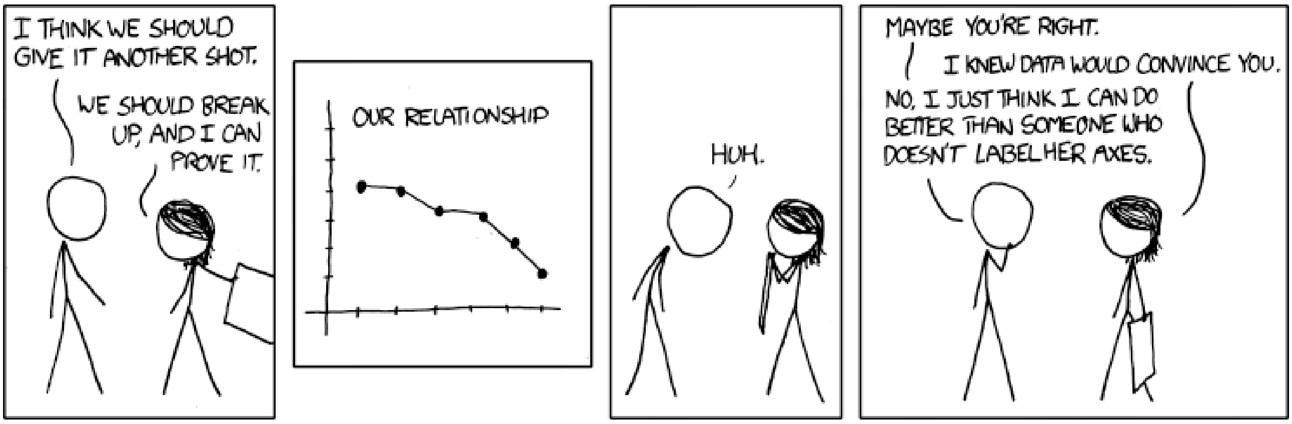


Figure 10.5.1. "Convincing" retrieved from http://imgs.xkcd.com/comics/convincing.png is used under a CC-BY-NC 2.5 licence

Bar Graphs

Bar graphs are commonly used to display and compare the average scores (means) of two or more groups or conditions. They are especially effective for visually highlighting differences between groups in a clear and straightforward way.

In an APA-style bar graph, like the one shown in Figure 10.5.2, each bar represents the mean score for a group or condition. These bars make it easy to compare groups at a glance. An additional important feature of bar graphs is the error bars, which are the small vertical lines extending upward and downward from the top of each bar.

Error bars indicate the variability of the data within each group. They typically represent the standard error of the mean rather than the standard deviation. The standard error is calculated by dividing the standard deviation of the group by the square root of the sample size. This measure is used because it provides an estimate of how much the group mean might vary from the true population mean.

Error bars are also helpful for assessing statistical significance. As a general rule, if the error bars of two groups do not overlap, it suggests that the difference between those groups is likely to be statistically significant. In other words, it is a visual cue that the observed difference is unlikely to have occurred by chance.

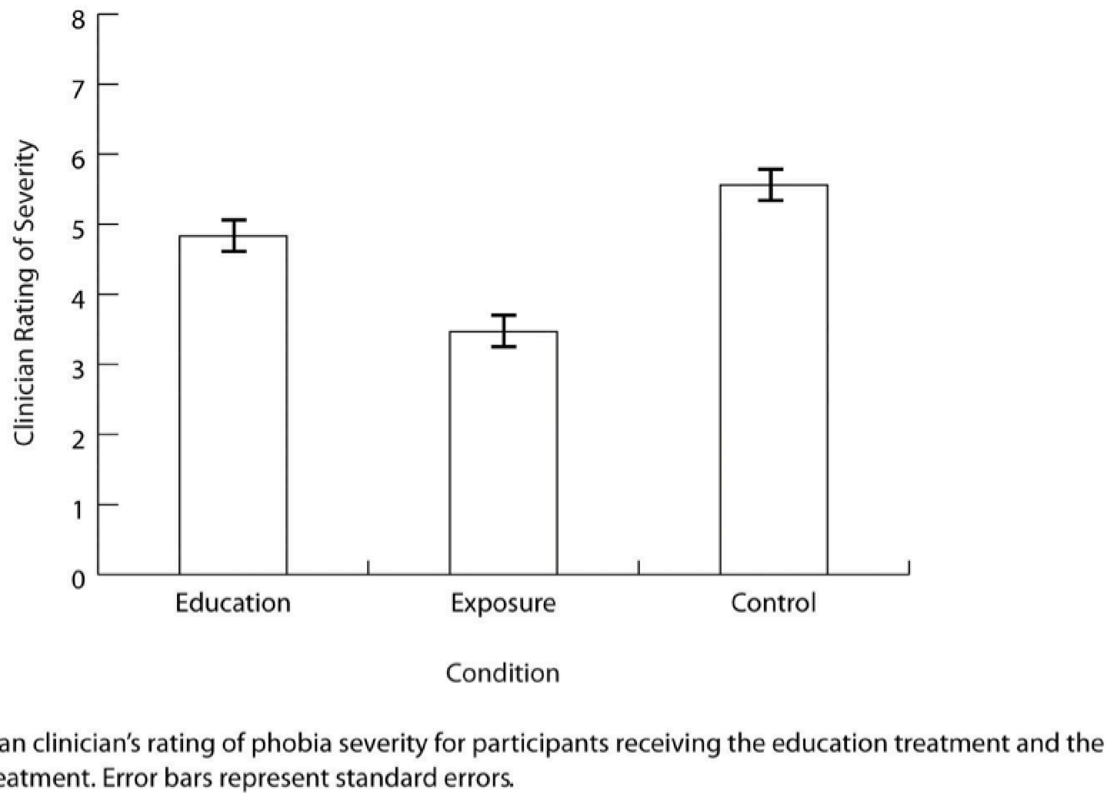


Figure 10.5.2. Sample APA-style bar graph, with error bars representing the standard errors, based on research by Ollendick and Colleagues by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Line Graphs

Line graphs are ideal for displaying data when the independent variable is continuous, for example, when it represents time or another variable measured on a numeric scale. They are also useful for showing correlations between quantitative variables when the independent variable has only a small number of distinct levels.

In a line graph, each point represents the average (mean) score on the dependent variable for participants at a specific level of the independent variable. These points are then connected by a line to visually demonstrate trends or patterns in the data. In Figure 10.5.3, an APA-style example of a line graph, error bars are included to show the standard error of the mean for each data point, adhering to APA formatting guidelines.

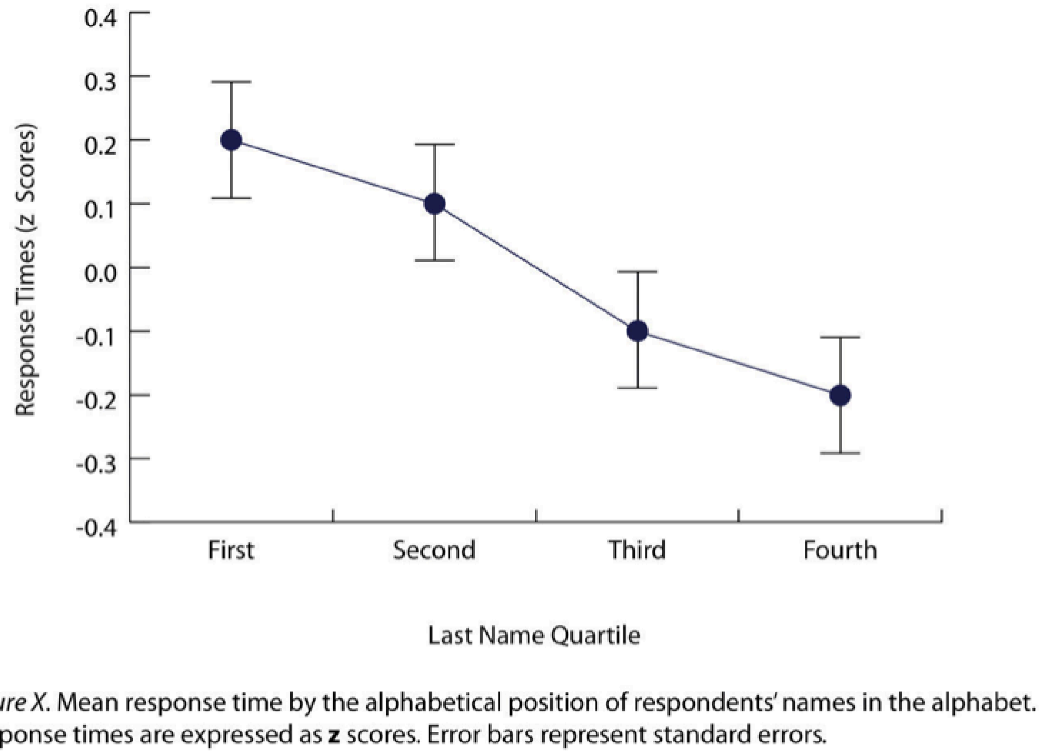


Figure 10.5.3. Sample APA-style line graph based on research by Carlson and Conard by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Interestingly, the data presented in a line graph could often be shown in a bar graph instead. For instance, in Figure 10.5.3, each point could be replaced with a bar reaching the same height, and the error bars would remain in place. This similarity highlights that both bar graphs and line graphs represent differences in average scores across levels of an independent variable.

However, there is a general convention in research for deciding which type of graph to use. Bar graphs are typically used when the x-axis represents categorical variables (e.g., treatment groups or conditions). In contrast, line graphs are preferred when the x-axis represents a quantitative variable (e.g., time intervals, dosage levels, or score ranges).

Scatterplots

Scatterplots are an effective way to show correlations and relationships between two quantitative variables, especially when the variable on the x-axis has a wide range of values. Unlike bar or line graphs, each point in a scatterplot represents one individual data point rather than an average or group mean. The points are not connected by lines, which helps to highlight the overall pattern or trend in the data.

In Figure 10.5.4, an APA-style scatterplot example, several important features are demonstrated. First, when both variables on the x-axis and y-axis are conceptually similar and measured on the same scale, like repeated measurements of self-esteem on two different occasions, the axes can be made equal in length to visually emphasise their similarity.

Second, sometimes multiple data points might overlap because two or more individuals have identical scores on both variables. This overlap can be addressed in a few ways:

- Offsetting the points slightly along the x-axis so they appear side by side.
- Adding a number in parentheses next to the point to indicate how many individuals share that score.
- Adjusting the size or darkness of the point to reflect the number of overlapping data points.

Finally, a scatterplot often includes a regression line. This is a straight line that best represents the overall trend of the data points. The regression line helps to show whether the relationship between the two variables is positive, negative, or neutral, and how closely the points align with this trend.

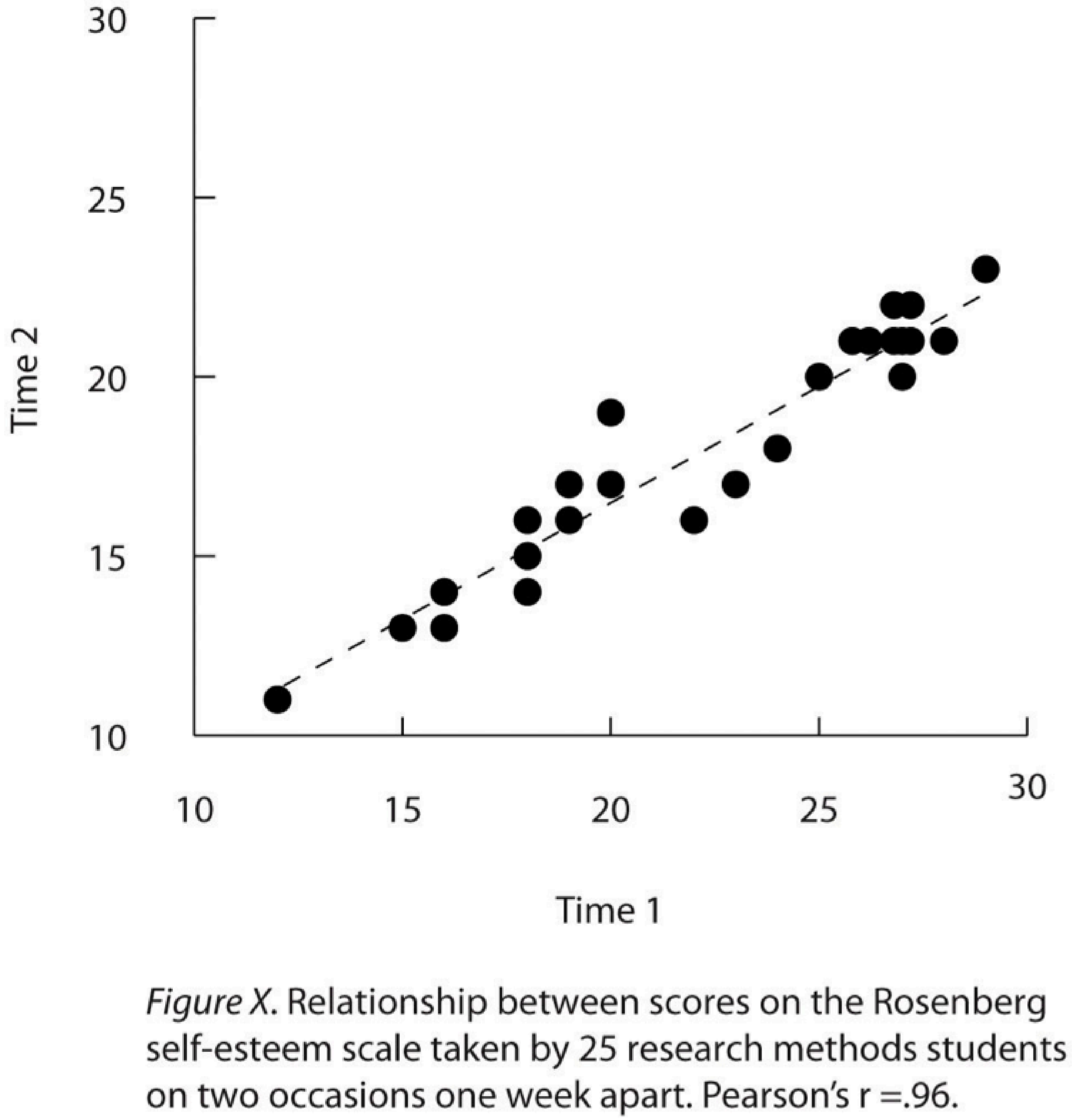


Figure 10.5.4. Sample APA-style scatterplot by R. S. Jhangiani et al. in [Research Methods in Psychology 4e](#) is used under a [CC BY-NC-SA licence](#)

Expressing Descriptive Statistics in Tables

Tables are a powerful tool for presenting large amounts of statistical data in a clear and organised way. Like graphs, they should serve a clear purpose, add valuable information, and be easy to understand on their own without requiring extensive reference to the text.

One of the most common uses of tables is to display multiple means and standard deviations from complex research designs involving several independent and dependent variables. For example, Figure

10.5.5 illustrates results from a hypothetical study similar to the one conducted by MacDonald and Martineau (2002). In their study, participants were categorised based on their self-esteem levels (high or low), placed in positive or negative moods, and then evaluated on their intentions toward unprotected sex as well as their attitudes toward unprotected sex.

The table in Figure 10.5.5 is structured clearly. Horizontal lines run across the top, bottom, and beneath column headings to improve readability. Each column is labelled appropriately, including the leftmost column, which provides context for the rows. Additionally, some column headings span multiple columns, allowing related data to be grouped together efficiently. APA-style tables are also numbered consecutively (e.g., Table 1, Table 2) and are accompanied by a concise yet descriptive title that summarises the table's purpose.

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 10.5.5. Sample APA-style table presenting means and standard deviations by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Another frequent use of tables is to present correlations among multiple variables, usually represented by Pearson's *r*. This type of table is called a correlation matrix. Figure 10.5.6 shows a correlation matrix based on a study by McCabe et al. (2010). The researchers examined the relationships between working memory and several cognitive abilities, such as executive function, vocabulary, and age.

From the table, we can quickly identify patterns. For example, the correlation between working memory and executive function was extremely strong at 0.96, while the correlation between working memory and vocabulary was more moderate at 0.27. Additionally, it is clear that most measures, except vocabulary, tend to decline with age.

One important feature of correlation matrices is that only half the table is filled in because the other half would simply duplicate the same values. For example, the correlation between working memory and age in the upper right corner is identical to the one between age and working memory in the lower left corner. Similarly, a variable's correlation with itself is always 1.00, which is typically replaced with dashes (—) to reduce clutter and improve readability.

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 10.5.6. Sample APA-style table (correlation matrix) based on research by McCabe and Colleagues by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

When presenting data in tables, it is important to remember that specific numerical results do not need to be repeated in the text. Instead, the text should focus on highlighting major trends and drawing attention to key details or patterns that are especially significant or relevant to the research question.

References

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. III, 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

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

10.6. CONDUCTING YOUR ANALYSES

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Analysing data can be a challenging task, even when you have a solid understanding of the statistical methods involved. Typically, you will be working with data collected from multiple participants, covering several variables. These might include demographic details like age and sex, independent and dependent variables, and possibly manipulation checks to verify experimental conditions.

The raw data you collect might come in various forms, including paper questionnaires, digital files filled with numbers or text, video recordings, or written observations. These different sources of information often need to be organised, coded, or merged into a cohesive dataset before analysis can begin. Additionally, you may encounter missing data, errors, or responses that seem unusual or inconsistent, all of which need to be addressed carefully.

In this section, we will explore practical strategies to streamline your data analysis process. By staying organised and approaching the task systematically, you can reduce errors, save time, and ensure your results are accurate and reliable.

Preparing Your Data for Analysis

Before analysing your data, whether it is in paper form or stored in a digital file, there are some essential steps to follow to ensure everything is organised, secure, and ready for processing.

First, make sure your data does not include any information that could identify individual participants. Confidentiality is crucial, so store raw data securely, either in a locked room or on a password-protected computer. Consent forms should be stored separately in another secure location. Additionally, create backup copies of your data, either photocopies or digital backups, and store them securely in a different location. Professional researchers typically keep these records for several years in case questions arise later about the data, procedure, or consent process.

Next, carefully review your raw data for completeness and accuracy. Check that all responses are legible, recorded correctly, and make sense. You might encounter missing responses, unclear answers, or obvious errors (e.g., someone marking "12" on a 1-to-10 scale). If these issues affect critical independent or dependent variables, or if too many responses are missing or questionable, you may need to exclude that participant's data from your analysis. However, do not delete or discard excluded data. Instead, set them aside, document the reasons for exclusion, and keep detailed notes as you will need to report this in your final analysis.

Once your data are clean and ready, you can enter them into a spreadsheet or statistical software like Microsoft Excel or SPSS. If your data are already in a digital file, ensure they are properly formatted for analysis. Typically, data are organised so that each row represents one participant, and each column represents one variable, with clear variable names at the top of each column.

For example, a typical data file shown in Table 10.6.1 starts with a column for participant identification numbers, followed by demographic variables (e.g., sex and age), independent variables (e.g., mood), multiple survey items (e.g., self-esteem questions), and dependent variables (e.g., intentions and attitudes). Categorical variables can be entered either as labels (e.g., "M" for male, "F" for female) or as numbers (e.g., "0" for negative mood and "1" for positive mood). While labels are more intuitive for reading, certain statistical analyses may require numerical coding. Tools like SPSS allow you to enter numerical values and attach corresponding labels for clarity.

Table 10.6.1. Sample data file

ID	SEX	AGE	MOOD	SE1	SE2	SE3	SE4	TOTAL	INT	ATT
1	M	20	1	2	3	2	3	10	6	5
2	F	22	1	1	0	2	1	4	4	4
3	F	19	0	2	2	2	2	8	2	3
4	F	24	0	3	3	2	3	11	5	6

If you are working with multiple-response measures, such as several survey items assessing self-esteem, it is better to enter each response as a separate variable in your spreadsheet rather than manually calculating a total score beforehand. Software tools like Excel or SPSS have built-in functions (e.g., "AVERAGE" in Excel or "Compute" in SPSS) to combine these responses accurately. This method reduces errors, allows you to check internal consistency, and provides flexibility if you decide to analyse individual survey items later.

Preliminary Analyses

Before diving into your primary research questions, it is important to run a few preliminary analyses to ensure your data are reliable and ready for deeper examination.

If you are using a multiple-response measure, start by checking its internal consistency. This ensures that the items on your measure are reliably capturing the same underlying concept. Statistical programs like

SPSS can calculate reliability coefficients such as Cronbach's α or Cohen's κ. If these seem too complex, you can still assess reliability with a simpler method like a split-half correlation, which compares how two halves of the measure align with each other.

Next, analyse each key variable on its own (this step is not necessary for manipulated independent variables since their values are determined by the researcher). Start by creating histograms for each variable to visualise their distributions. Pay attention to their shapes and calculate common measures of central tendency (e.g., mean, median, mode) and variability (e.g., standard deviation). Be sure you understand what these statistics reveal about your data. For example, if participants rated their happiness on a 1-to-10 scale and the distribution is unimodal and negatively skewed, with a mean of 8.25 and a standard deviation of 1.14, it means most participants rated themselves fairly high on happiness, with a few giving noticeably lower ratings.

At this stage, it is also essential to identify outliers. These are data points that stand out as extreme compared to the rest of your dataset. Investigate these outliers carefully to determine whether they result from simple data-entry errors. If you find a mistake, correct it and move on. However, if an outlier seems to stem from a misunderstanding or lack of effort from a participant, you might need to consider excluding it. For example, in a reaction-time study where most participants responded within a few seconds, a response time of three minutes would likely indicate confusion or inattention. Including such an extreme value would significantly distort the mean and standard deviation.

If you decide to exclude outliers, document your reasons carefully and apply the same criteria consistently across all participants. Keep detailed notes on which data points were removed, why they were excluded, and the rules you followed. When you report your results, make sure to mention how many participants or responses were excluded and the criteria used for exclusion. Importantly, do not delete or discard excluded data. Set them aside in case you or another researcher needs to review them later.

It is worth noting that not all outliers are errors or misunderstandings. Sometimes, they genuinely reflect extreme but valid responses. For example, in a survey on the number of sexual partners among university students, most participants might report fewer than 15, but a handful might report 60 or 70. While these numbers could be errors, exaggerations, or misunderstandings, they might also be accurate reflections of those participants' experiences.

In such cases, there are a few strategies you can use. One approach is to rely on statistics like the median, which are less affected by extreme values. Another approach is to run your analysis twice, once with the outliers included and once without them. If the results are essentially the same, it is usually safe to leave the outliers in the dataset. If the results differ significantly, you can report both analyses and explain how the outliers influenced the findings.

Planned and Exploratory Analyses

Once your data are prepared and preliminary analyses are complete, you are ready to address your primary research questions. When designing your study, you likely had specific hypotheses in mind, such as predictions about relationships or patterns you expected to find in the data. Testing these predictions involves planned analyses, where you focus on analysing the relationships you anticipated.

For example, if your hypothesis predicted a difference between group or condition means, you would calculate the means and standard deviations for each group, create a bar graph to visualise the results, and calculate Cohen's *d* to measure the size of the difference. If your hypothesis involved a correlation between two quantitative variables, you would create a scatterplot or line graph (making sure to check for any signs of nonlinearity or restriction of range) and calculate Pearson's *r* to measure the strength and direction of the relationship.

After completing your planned analyses, you might decide to look for additional patterns or relationships in your data that you did not predict beforehand. These are called exploratory analyses because they are not based on pre-existing hypotheses. Exploratory analyses can uncover unexpected findings that might inspire future research or provide valuable insights for the discussion section of your report.

As psychologist Daryl Bem (2003) suggests, exploratory analysis often involves examining the data from multiple perspectives. You might analyse subgroups separately (e.g., by sex), create new composite scores by combining variables, or reorganise the data in different ways to reveal potential patterns. If an interesting trend emerges, you might explore whether similar evidence exists elsewhere in your dataset. While Bem humorously describes this as a "fishing expedition", the goal is to uncover meaningful insights hidden in the data.

However, it is important to distinguish planned analyses from exploratory analyses when presenting your results. Planned analyses are based on specific hypotheses and have a clearer foundation, while exploratory analyses are more open-ended and carry a higher risk of identifying patterns that occurred purely by chance. This risk is known as a Type 1 error, where a random anomaly is mistaken for a genuine finding.

Because of this risk, findings from exploratory analyses should be interpreted cautiously and ideally tested again in a follow-up study before being presented as reliable results. In your report, make it clear which results came from planned analyses and which emerged during exploratory analysis. If you discover intriguing patterns during exploratory analysis, describe them as potential areas for further investigation rather than definitive conclusions.

Understanding Your Descriptive Statistics

Before diving into inferential statistics, which help determine whether your study's results are likely to apply to the larger population, it is essential to fully understand your descriptive statistics. These statistics tell the story of what actually happened in your study, providing a clear snapshot of your data.

For example, imagine a study where a treatment group of 50 participants has an average score of 34.32 with a standard deviation (*SD*) of 10.45, while a control group of 50 participants has an average score of 21.45 with an *SD* of 9.22. Additionally, the effect size (Cohen's *d*) is a very strong 1.31. Even without running a formal inferential statistical test like a *t*-test, it is already clear from these descriptive statistics that the treatment had a significant impact.

Similarly, consider a scatterplot showing a random cloud of data points and a Pearson's *r* value of −0.02. This tiny correlation tells you that there is essentially no relationship between the two variables. Again, while inferential statistical testing would still be part of a formal report, the descriptive statistics alone already paint a clear picture.

References

Bem, D. J. (1987). *Writing the empirical journal article*. Psychology Press.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

10.7. SUMMARY

By Marc Chao

Summary

A variable's distribution reveals how its values are spread across categories or levels, offering insights into patterns such as frequency, range, and outliers. Frequency tables provide a structured way to display distributions, with grouped tables simplifying large datasets by combining values into intervals. Visual tools like histograms complement these tables, showcasing distribution shapes such as unimodal, bimodal, symmetrical, or skewed, which highlight central tendencies, clustering, and potential outliers. Understanding these characteristics is crucial for accurate data interpretation and meaningful trend identification.

Central tendency represents the central value in a dataset, with the mean, median, and mode serving as its primary measures. The mean, calculated as the average, is widely used but sensitive to outliers. The median, dividing data into equal halves, is more robust against extreme values and ideal for skewed distributions. The mode, identifying the most frequent value, applies to both numerical and categorical data. While these measures often align in symmetrical distributions, they diverge in skewed datasets, requiring careful selection based on the data's characteristics. Using multiple measures often provides a richer understanding of a dataset's central tendencies.

Variability captures the spread of data values around the centre, complementing measures of central tendency. The range, as the simplest measure, is easily affected by outliers, while the standard deviation provides a more precise depiction of average deviations from the mean. Variance, an intermediate calculation for standard deviation, is essential in advanced statistical methods but less interpretable. Tools like percentile ranks and *z*-scores further contextualise individual scores, indicating relative positions within a dataset and enabling comparisons across datasets. These measures deepen our understanding of data distribution and support meaningful statistical interpretation.

Psychological research often investigates relationships between variables, focusing on differences between groups or conditions and correlations between quantitative variables. Group comparisons use means, standard deviations, and effect sizes like Cohen's *d* to assess disparities, while scatterplots and Pearson's *r* measure and visualise correlations. Researchers must account for nonlinear relationships and restricted ranges, which can obscure or distort findings.

After analysing data with descriptive statistics, researchers communicate their findings through text, figures, and tables, adhering to APA guidelines for clarity and consistency. Descriptive statistics in writing balance precision and readability, using proper formatting for values like means (*M*) and standard deviations (*SD*). Figures, including bar graphs, line graphs, and scatterplots, visually highlight trends and relationships, with APA standards emphasising simplicity, clear labelling, and the inclusion of error bars where necessary. Tables effectively present complex datasets, such as means or correlations, with concise titles and structured layouts that enhance understanding. The combination of text, visuals, and tables ensures findings are communicated effectively and avoids redundancy.

Data analysis begins with organising, cleaning, and preparing data from multiple variables, ensuring confidentiality, addressing errors, and structuring datasets for analysis. Preliminary steps include evaluating internal consistency, visualising distributions, and addressing outliers to ensure data accuracy. Planned analyses test hypotheses through comparisons and correlations, while exploratory analyses examine unexpected patterns, requiring cautious interpretation due to the risk of random anomalies. Descriptive statistics, including means, standard deviations, and effect sizes, offer a foundational understanding of data trends.

CHAPTER 11: INFERENTIAL STATISTICS

In psychological research, the ultimate aim is to go beyond merely describing data from a sample. Researchers strive to make meaningful inferences about a larger population based on the sample data they collect. While descriptive statistics such as means, medians, and correlation coefficients are valuable for summarising sample data, they are often just the first step. The real challenge lies in determining whether the patterns observed in a sample reflect true phenomena in the population or are merely the result of random chance. This is where inferential statistics plays a crucial role.

Imagine a researcher studying depression symptoms among 50 adults diagnosed with clinical depression. Calculating the average number of symptoms in this sample provides a snapshot of the group. However, the researcher's goal is not limited to this specific sample; they aim to understand the broader population of individuals with clinical depression. This process of making inferences about population parameters based on sample statistics is central to inferential statistics.

Yet, drawing conclusions about a population from a sample is not without challenges. Sampling error, which refers to the random variability inherent in selecting a subset of individuals, can result in fluctuating sample statistics. For instance, the average number of depressive symptoms might be 8.73 in one sample, 6.45 in another, and 9.44 in a third. Similarly, correlations between two variables could vary from +0.24 to -0.04 to +0.15 across samples. These variations, stemming from random chance, complicate the task of discerning whether observed relationships or differences in a sample truly exist in the population or are mere artifacts of sampling error.

To address this challenge, researchers turn to null hypothesis testing, a cornerstone of inferential statistics. This systematic approach helps distinguish between two possibilities for any observed statistical relationship:

1. The relationship exists in the population, and the sample reflects this genuine pattern.
2. The relationship is a product of sampling error, with no real effect in the population.

By applying statistical techniques, researchers can assess the likelihood of their findings arising by chance under the assumption that the null hypothesis is true. This chapter delves into the principles and applications of inferential statistics, with a particular focus on null hypothesis testing, its underlying logic, and the tools researchers use to draw conclusions about populations. We will also explore common misconceptions, limitations, and alternative methods, providing a comprehensive understanding of this critical aspect of research methodology.

Learning Objectives

By the end of this chapter, you should be able to:

- **Understand statistical significance and the *p-value*:** Define the *p*-value and explain its correct interpretation as the probability of obtaining an observed result if the null hypothesis is true.
- **Critically evaluate the use of *p-values*:** Assess the limitations of relying on a rigid $p < 0.05$ threshold to determine significance and understand how this practice can lead to arbitrary distinctions between "significant" and "non-significant" results.
- **Incorporate effect sizes and confidence intervals:** Explain the importance of reporting effect sizes to convey the strength of a relationship and use confidence intervals to provide a range of plausible values for population parameters.
- **Address the file drawer problem and *p*-hacking:** Understand how practices like selective reporting and *p*-hacking contribute to distorted research findings. Learn how transparency measures, such as registered reports and sharing non-significant results, can improve the reliability of scientific research.
- **Explore alternatives to null hypothesis testing:** Recognise the limitations of null hypothesis testing and explore alternative approaches like Bayesian statistics, which provide more nuanced insights by updating probabilities based on observed data.

11.1. NULL HYPOTHESIS TESTING

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

In psychological research, scientists often measure one or more variables in a sample and then calculate descriptive statistics, such as means or correlation coefficients, to summarise the data. These statistics provide useful insights about the sample, but the ultimate goal of most research is to make conclusions about the larger population from which the sample was drawn. In statistics, these population values are called parameters.

For example, imagine a researcher measures the number of depressive symptoms in 50 adults diagnosed with clinical depression and calculates the average number of symptoms in that sample. The researcher is not just interested in the average for that specific group but wants to make an inference about the average number of depressive symptoms across all adults with clinical depression.

However, sample statistics are imperfect estimates of population parameters because of random variability, a phenomenon known as sampling error. Even if samples are selected randomly from the same population, the statistics can vary. For instance, in one sample, the average number of depressive symptoms might be 8.73, while in another it might be 6.45, and in a third, it could be 9.44. Similarly, the correlation between two variables might show values like +0.24, -0.04, or +0.15 across different samples. This variability is normal and expected, and it does not mean anyone made a mistake, it is simply the result of random chance.

This variability creates a challenge: when researchers observe a statistical relationship in a sample, it is not always clear whether the relationship truly exists in the population or if it is just a result of sampling error.

For example, a small difference between two group means in a sample might suggest a real difference in the population. However, it could also mean that no real difference exists, and the observed difference is merely due to random variation. Similarly, a correlation value of -0.29 in a sample might indicate a negative relationship in the population, or it might just be noise caused by sampling error.

In essence, every statistical relationship in a sample has two possible explanations:

- The relationship exists in the population, and the sample reflects this real relationship.
- There is no real relationship in the population, and the sample result is simply due to sampling error.

The purpose of null hypothesis testing is to help researchers distinguish between these two possibilities and make informed conclusions about whether the patterns they see in their data are likely to reflect real relationships in the population.

The Logic of Null Hypothesis Testing

Null hypothesis testing (often called null hypothesis significance testing or NHST) is a statistical method used to determine whether a relationship observed in a sample reflects a real relationship in the population or if it simply occurred by chance.

At the core of NHST are two competing explanations: the null hypothesis (H_0) and the alternative hypothesis (H_1). The null hypothesis suggests that there is no real relationship in the population and that any observed relationship in the sample is due to random chance or sampling error. In simpler terms, it assumes that the sample result is just a coincidence. The alternative hypothesis, on the other hand, proposes that there is a real relationship in the population and that the sample result reflects this genuine relationship.

Every statistical result in a sample can be interpreted in one of these two ways: either it happened by chance, or it represents a real relationship in the population. Researchers need a systematic way to decide between these two interpretations, and NHST provides that framework.

The process follows a clear logic. First, researchers assume the null hypothesis is true, meaning they start from the assumption that there is no relationship in the population. Then they calculate how likely the sample result (or one even more extreme) would be if the null hypothesis were true. If the result is extremely unlikely under the null hypothesis, researchers reject the null hypothesis in favour of the alternative hypothesis. If the result is not extremely unlikely, they fail to reject the null hypothesis and conclude that there is not enough evidence to claim a real relationship exists.

To illustrate, consider the study by Mehl and colleagues, who investigated differences in talkativeness between men and women. They asked, "If there were no real difference in talkativeness in the population, how likely would it be to observe a difference of $d = 0.06$ in our sample?" Their analysis showed that such a small difference was fairly likely under the null hypothesis. As a result, they failed to reject the null hypothesis and concluded there was no evidence of a meaningful difference in the population.

In contrast, Kanner and colleagues examined the correlation between daily hassles and symptoms. They asked, "If there were no real correlation in the population, how likely would it be to observe a strong correlation of +0.60 in our sample?" Their analysis indicated that such a strong correlation would be very unlikely under the null hypothesis. Therefore, they rejected the null hypothesis and concluded that there is a positive correlation between these variables in the population.

A key step in NHST is determining the *p*-value, which represents the probability of obtaining a sample result (or one even more extreme) if the null hypothesis were true. A low *p*-value indicates that the result is unlikely under the null hypothesis, leading researchers to reject the null hypothesis. A high *p*-value, on the other hand, suggests that the result could reasonably occur by chance, so researchers fail to reject the null hypothesis.

But how low must a *p*-value be to reject the null hypothesis? Researchers typically use a threshold called alpha (α), which is almost always set at 0.05. If the *p*-value is 0.05 or lower, it means there is less than a 5% chance of obtaining such an extreme result if the null hypothesis were true. In this case, the result is considered statistically significant, and the null hypothesis is rejected. If the *p*-value is greater than 0.05, the null hypothesis is not rejected, but this does not mean it is accepted as true, only that there is not enough evidence to reject it.

To avoid confusion, researchers avoid saying they "accept" the null hypothesis. Instead, they use the phrase "fail to reject the null hypothesis", emphasising that the evidence simply was not strong enough to support rejecting it.

Understanding these principles is crucial for interpreting statistical results accurately and avoiding common misunderstandings about what a *p*-value really represents.

The Misunderstood *p*-Value

The *p*-value is one of the most frequently misunderstood concepts in psychological research, even among experienced researchers. Misinterpretations of the *p*-value are so common that they sometimes appear in statistics textbooks as well.

A widespread mistake is thinking that the *p*-value represents the probability that the null hypothesis is true or the likelihood that the sample result happened purely by chance (Figure 11.1.1). For example, someone might incorrectly conclude that a *p*-value of 0.02 means there is only a 2% chance that the result was due to random chance and a 98% chance that the observed relationship is real. This interpretation is wrong.

In reality, the *p*-value indicates the probability of obtaining a result as extreme as the one observed (or more extreme) if the null hypothesis were true. A *p*-value of 0.02 means that if there were truly no relationship in the population (if the null hypothesis were correct), a sample result this extreme would occur only 2% of the time due to random chance.

To avoid this common misunderstanding, it is essential to remember that the *p*-value does not tell us the probability that the null hypothesis is true or false. Instead, it tells us how likely it is to observe the sample result (or something even more extreme) under the assumption that the null hypothesis is correct.

Figure 11.1.1. "Null Hypothesis" retrieved from http://imgs.xkcd.com/comics/null_hypothesis.png is used under a CC-BY-NC 2.5 licence

Role of Sample Size and Relationship Strength

Null hypothesis testing seeks to answer this question: "If the null hypothesis were true, what is the probability of observing a sample result as extreme as this one?" This probability is the *p*-value, and it depends on two key factors: the strength of the relationship and the sample size. Specifically, stronger relationships and larger samples make it less likely that the observed result would occur if the null hypothesis were true, leading to a lower *p*-value.

To clarify, imagine two scenarios:

1. **Large sample, strong relationship:** Suppose you compare a sample of 500 women and 500 men on a psychological characteristic, and Cohen's *d* is a strong 0.50. If there were truly no sex difference in the population, such a strong result from such a large sample would be very unlikely. The null hypothesis would likely be rejected.
2. **Small sample, weak relationship:** Now imagine a similar study comparing three women and three men, but Cohen's *d* is a weak 0.10. If there were no sex difference in the population, this weak result from such a small sample would be entirely plausible. The null hypothesis would likely be retained.

This explains why strong results in large samples are more likely to lead to rejecting the null hypothesis, while weak results in small samples are not.

However, the relationship between these two factors is not always so straightforward. A weak result can still be statistically significant if the sample is very large, and a strong relationship can be significant even with a small sample. This trade-off between relationship strength and sample size is summarised in Table 11.1.1, which provides a rough guideline for determining statistical significance.

Table 11.1.1 How relationship strength and sample size combine to determine whether a result Is statistically significant

Relationship strength			
Sample Size	Weak	Medium	Strong
Small (N = 20)	No	No	d = Maybe$r$ = Yes
Medium (N = 50)	No	Yes	Yes
Large (N = 100)	d = Yes$r$ = No	Yes	Yes
Extra large (N = 500)	Yes	Yes	Yes

From Table 11.1.1:

- **Weak relationships:** These are never statistically significant in small or medium samples but can be significant in very large samples.
- **Strong relationships:** These are always statistically significant with medium or larger samples and sometimes even with small samples, depending on specific factors.

This understanding is invaluable for developing intuitive judgement about statistical significance. For instance, if you observe a strong relationship with a medium sample, you should expect to reject the null hypothesis. If the formal test suggests otherwise, it signals a need to review your computations or interpretations.

Statistical Significance vs. Practical Significance

Statistical significance and practical significance are two different but equally important concepts in interpreting research results. A result can be statistically significant without being practically meaningful. This distinction is crucial because statistical significance only tells us that an observed effect is unlikely to have occurred by random chance, not whether the effect is strong or useful in a real-world context (Figure 11.1.2).

For example, Janet Shibley Hyde highlighted findings on sex differences in areas like mathematical problem-solving and leadership ability. While these differences are often statistically significant, they are usually so small that they have little practical relevance. Despite their statistical significance, these findings should not influence major decisions, such as which college courses to pursue or whom to vote for.

This distinction becomes even more important in fields like clinical psychology. A study might find that a new treatment for social phobia produces a statistically significant improvement in symptoms. However, if the improvement is very small, it might not justify the costs, time, and effort required to implement the treatment, especially if existing treatments deliver similar or better results.

In such cases, researchers refer to this concept as clinical significance, which focuses on whether the effect is meaningful or useful in a practical setting. Even if a study meets the statistical threshold for significance, it might lack real-world impact if the effect size is minimal or the costs outweigh the benefits.

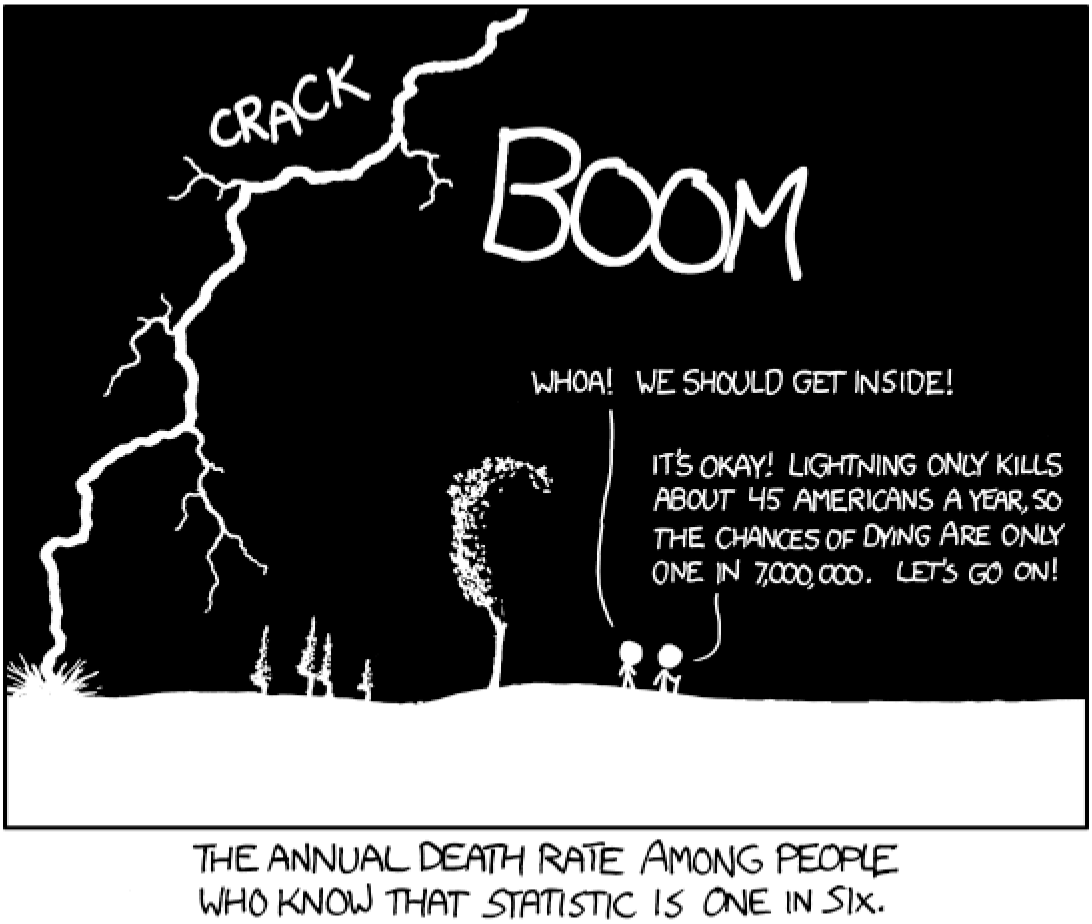


Figure 11.1.2. "Conditional Risk" retrieved from http://imgs.xkcd.com/comics/conditional_risk.png is used under a CC-BY-NC 2.5 licence

Chapter Attribution

Content adapted, with editorial changes, from:

11.2. THE T-TEST

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

In psychology, many studies examine differences between two means, and the *t*-test is the most common tool for analysing such relationships. This section focuses on three types of *t*-tests, each suited to different research designs:

1. **One-Sample *t*-Test:** Compares a sample mean to a known population mean.
2. **Dependent-Samples *t*-Test:** Used when comparing two related groups, such as the same participants tested at different times.
3. **Independent-Samples t-Test:** Compares the means of two independent groups, such as different sets of participants.

Even if you have taken a statistics course before, this section will help refresh your understanding of these essential tools.

One-Sample t-Test

The one-sample *t*-test is used to compare a sample mean (M) to a hypothetical population mean (μ_0) that serves as a reference point. The test evaluates two competing hypotheses:

- **Null Hypothesis (H_0):** The population mean (μ) equals the hypothetical mean (μ_0).
- **Alternative Hypothesis (*H*◈):** The population mean (μ) is different from the hypothetical mean (μ_0).

To decide between these hypotheses, the *t*-test calculates the probability (*p*-value) of obtaining a sample mean as extreme as the one observed if the null hypothesis were true. First, a statistic called t is calculated using this formula:

$$t = \frac{M - \mu_0}{\left(\frac{SD}{\sqrt{N}}\right)}$$

Here:

- M = Sample mean
- μ_0 = Hypothetical population mean
- SD = Sample standard deviation
- N = Sample size

The Role of the *t* Distribution

The *t*-statistic is useful because its behaviour under the null hypothesis is well understood as it follows a *t* distribution. This distribution is unimodal, symmetric, and centred at 0. Its shape depends on degrees of freedom (*df*), calculated as $N - 1$ for a one-sample *t*-test.

For example, with $df = 24$, the *t* distribution looks like Figure 11.2.1, where extreme *t* values occur in the tails. To calculate the *p*-value, the proportion of scores in the tails of this distribution that are as extreme as the observed *t* value is determined. For instance, a *t* value of 1.50 with $df = 24$ corresponds to a *p*-value of 0.14.

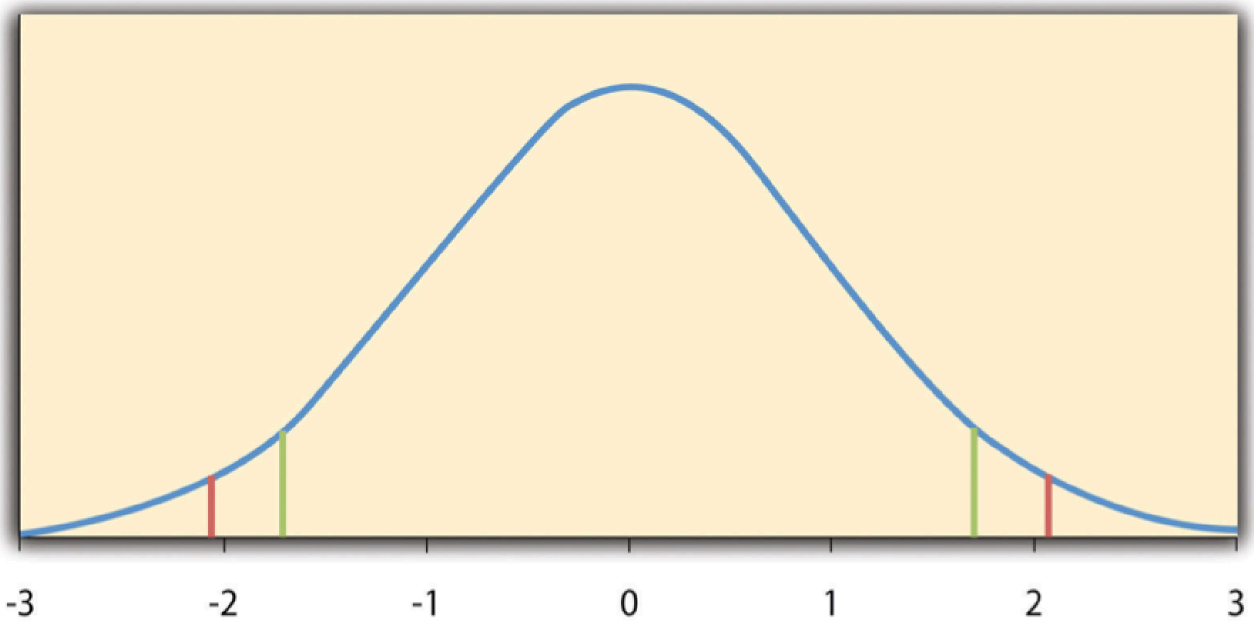


Figure 11.2.1. Distribution of t scores (with 24 degrees of freedom) when the null hypothesis is true. The red vertical lines represent the two-tailed critical values, and the green vertical lines the one-tailed critical values when α = .05, by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Decision Rule

Using statistical software, you can directly obtain the *t* value and *p*-value. Here's how you interpret the results:

- If $p \leq 0.05$, reject the null hypothesis and conclude that the population mean likely differs from the hypothetical mean.
- If $p > 0.05$, retain the null hypothesis, meaning there is insufficient evidence to conclude a difference.

When manually computing the *t* value, you can refer to a critical *t*-table (like Table 11.2.1). The table provides critical *t* values for different degrees of freedom (*df*) and significance levels ($\alpha = 0.05$). For a two-tailed test, any *t* value beyond the critical values (e.g., ±2.064 for $df = 24$) leads to rejecting the null hypothesis.

Table 11.2.1. Table of critical values of t when α = .05

df	One-tailed Critical value	Two-tailed
3	2.353	3.182
4	2.132	2.776
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.812	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145
15	1.753	2.131
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
30	1.697	2.042
35	1.690	2.030
40	1.684	2.021
45	1.679	2.014
50	1.676	2.009
60	1.671	2.000
70	1.667	1.994

df	**One-tailed**	**Two-tailed**
80	1.664	1.990
90	1.662	1.987
100	1.660	1.984

One-Tailed vs. Two-Tailed Tests

Two-Tailed Test:

- Used when you suspect the sample mean might differ from the hypothetical mean, but do not know in which direction.
- Reject the null hypothesis if the t value falls in either tail of the distribution (e.g., below −2.064 or above +2.064 for $df = 24$).

One-Tailed Test:

- Applied when you expect a difference in a specific direction.
- You decide beforehand whether to test for a sample mean higher or lower than the hypothetical mean.
- For $df = 24$, one-tailed critical values are less extreme (±1.711). However, the disadvantage is that you lose the ability to detect differences in the opposite direction.

In both cases, the significance level ($\alpha = 0.05$) is maintained. The one-tailed test gives more power to detect a directional difference but completely ignores the possibility of a difference in the opposite direction.

Example: One-Sample t-Test

A health psychologist is interested in evaluating how accurately university students estimate the calorie content of a chocolate chip cookie. The cookie contains 250 calories, which serves as the hypothetical population mean (μ_0). The psychologist's null hypothesis (H_0) states that the population mean of students' calorie estimates (μ) is equal to 250. Since he has no prior expectation about whether students overestimate or underestimate, he decides to conduct a two-tailed test to examine deviations in either direction.

To test the hypothesis, the psychologist gathers calorie estimates from a sample of 10 university students. Their individual estimates are as follows:

- 250, 280, 200, 150, 175, 200, 200, 220, 180, 250

From this data, the psychologist calculates the following sample statistics:

- **Sample Mean (*M*):** 212.00 calories
- **Sample Standard Deviation (*SD*)**: 39.17 calories
- **Sample Size (*N*):** 10

The psychologist uses the one-sample *t*-test to determine if the sample mean significantly differs from the hypothetical population mean. The formula for the *t*-test is:

$$t = \frac{M - \mu_0}{\left(\frac{SD}{\sqrt{N}}\right)}$$

Where:

- M = Sample mean = 212.00
- μ_0 = Hypothetical population mean = 250
- SD = Sample standard deviation = 39.17
- N = Sample size = 10

Substituting the values into the formula:

$$t = \frac{212 - 250}{\frac{39.17}{\sqrt{10}}} = \frac{-38}{12.39} = -3.07$$

The computed t value is −3.07.

The psychologist enters the data into statistical software (e.g., SPSS or Excel). The software calculates the p value for this t-score with $df = N - 1 = 10 - 1 = 9$. The output shows a two-tailed p value of 0.013.

- **Decision:** Since $p = 0.013$ is less than the significance level $\alpha = 0.05$, the psychologist rejects the null hypothesis.
- **Conclusion:** There is sufficient evidence to conclude that university students significantly underestimate the calorie content of a chocolate chip cookie.

If performing the test manually, the psychologist would consult a table of critical t values (e.g., Table 11.2.1). For a two-tailed test with $df = 9$ and $\alpha = 0.05$, the critical value of t is:

$t_{critical} = \pm 2.262$

- The observed t value ($t = -3.07$) is more extreme than the critical value (−2.262), leading to the same conclusion: the null hypothesis is rejected.

The psychologist reports the results as follows:

- $t(9) = -3.07, p = .01$

Key formatting points in APA style include:

- The symbols t and p are italicised.
- The degrees of freedom (df) are placed in parentheses immediately after t.
- Both the t value and p value are rounded to two decimal places.

If the psychologist had a strong reason to believe that students would specifically underestimate the calories, a one-tailed test could be used. In this case:

- The critical value for $df = 9$ at $\alpha = 0.05$ would be -1.833 (for the lower tail).
- Because $t = -3.07$ is more extreme than -1.833, the null hypothesis would still be rejected.

However, one-tailed tests have limitations. If the data showed an overestimation (a positive t value), the null hypothesis could not be rejected, as one-tailed tests only examine deviations in a pre-specified direction.

Dependent-Samples *t*-Test

The dependent-samples *t*-test, also called the paired-samples *t*-test, is used when comparing two means from the same group of participants. This test is typically applied in scenarios like pretest-posttest designs or within-subjects experiments, where each participant is tested under two conditions or at two different times.

Key Hypotheses

- **Null hypothesis (H_0):** The population means for the two conditions or time points are the same ($\mu_0 = 0$).
- **Alternative hypothesis ($H_⟐$):** The population means for the two conditions or time points are different ($\mu_0 \neq 0$).

If there is a clear expectation about the direction of the difference (e.g., an increase or decrease), a one-tailed test can be used instead of the more common two-tailed test.

How the Dependent-Samples *t*-Test Works

Think of the dependent-samples *t*-test as a variation of the one-sample *t*-test, with one additional step:

Calculate the difference scores:

- For each participant, compute the difference between their two scores (e.g., posttest score minus pretest score). These difference scores summarise the change for each individual.

Apply the One-Sample *t*-test:

- Once the difference scores are calculated, the test becomes a one-sample *t*-test conducted on the mean of the difference scores.
- The hypothetical population mean (μ_0) for the difference scores is 0 because, under the null hypothesis, there is no average difference between the two conditions or time points.

Reframing the Hypotheses

- Under the null hypothesis, the mean difference score in the population is 0 ($\mu_0 = 0$).
- The alternative hypothesis states that the mean difference score in the population is not 0 ($\mu_0 \neq 0$).

This approach allows the dependent-samples *t*-test to determine whether the observed differences between the two means are statistically significant, reflecting a real effect rather than random variation.

Example: Dependent-Samples *t*-Test

A health psychologist seeks to determine whether a specialised training program can enhance people's ability to estimate the calorie content of junk food accurately. To test the program's effectiveness, a pretest-posttest study is conducted with a sample of 10 participants. Each participant provides an estimate of the calories in a chocolate chip cookie before and after completing the training program. Since the psychologist expects the program to improve calorie estimates, a one-tailed test is used to analyse the data.

Pretest and Posttest Data Collection

The calorie estimates provided by the participants are recorded as follows:

- **Pretest estimates:** 230, 250, 280, 175, 150, 200, 180, 210, 220, 190
- **Posttest estimates:** 250, 260, 250, 200, 160, 200, 200, 180, 230, 240

The goal is to evaluate whether the posttest estimates, on average, are higher than the pretest estimates, indicating an improvement in accuracy due to the training.

To assess the change in estimates for each participant, the difference scores are calculated by subtracting the pretest estimates from the posttest estimates. A positive difference score indicates that a participant's calorie estimate increased after the training, while a negative score indicates a decrease.

- Difference Score = Posttest Estimate − Pretest Estimate

The resulting difference scores are:

- **Difference scores:** 20, 10, −30, 25, 10, 0, 20, −30, 10, 50

The direction of subtraction (posttest minus pretest) does not matter as long as it is applied consistently for all participants.

From the difference scores, the following descriptive statistics are calculated:

- **Mean of the difference scores (*M*):** 8.50
- **Standard deviation of the difference scores (*SD*):** 27.27
- **Sample size (*N*):** 10

The null hypothesis (H_0) states that the mean difference in calorie estimates after training is zero ($\mu_0 = 0$), implying no effect of the training program. The alternative hypothesis (H_A) posits that the mean difference is greater than zero ($\mu > 0$), consistent with an expected improvement.

The *t*-statistic is calculated using the following formula:

$$t = \frac{M - \mu_0}{\left(\frac{SD}{\sqrt{N}}\right)}$$

Where:

- M = Mean of the difference scores = 8.50
- μ_0 = Hypothetical population mean = 0
- SD = Standard deviation of the difference scores = 27.27
- N = Sample size = 10

Substituting the values into the formula:

$$t = \frac{8.5 - 0}{27.27/\sqrt{10}} = 1.11$$

The calculated *t* value is 1.11.

The data can be entered into software such as SPSS, Excel, or an online *t*-test calculator. For $df = N - 1 = 10 - 1 = 9$, the one-tailed *p* value is 0.148.

- **Decision:** Since $p = 0.148$ is greater than the significance threshold ($\alpha = 0.05$), the psychologist fails to reject the null hypothesis.
- **Conclusion:** There is insufficient evidence to conclude that the training program significantly improves calorie estimation accuracy.

If calculating by hand, the psychologist would refer to a critical *t* value table. For a one-tailed test with $df = 9$ and $\alpha = 0.05$, the critical *t* value is 1.833.

- **Comparison:** The observed *t* value ($t = 1.11$) is less extreme than the critical value ($t = 1.833$).
- **Decision:** The psychologist fails to reject the null hypothesis, consistent with the software output.

The results suggest that the training program does not lead to a statistically significant improvement in calorie estimation. While the mean difference score (8.50) indicates a slight increase in posttest estimates, this difference is not large enough to rule out the possibility that it occurred by chance.

The Independent-Samples *t*-Test

The independent-samples *t*-test is used to compare the means (M_1 and M_2) of two separate groups. These groups may have been exposed to different conditions in a between-subjects experiment or represent naturally occurring categories in a cross-sectional study (e.g., women vs. men, extraverts vs. introverts).

Hypotheses

- **Null hypothesis (H_0):** The population means are equal ($\mu_1 = \mu_2$).
- **Alternative hypothesis (H_A):** The population means are not equal ($\mu_1 \neq \mu_2$).

If there is strong evidence to expect a difference in a specific direction, a one-tailed test may be used instead of a two-tailed test.

Formula for the *t*-Statistic

The formula for the independent-samples *t*-test accounts for two sample means, their variances, and sample sizes. The *t*-statistic is calculated as:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD^2{}_1}{n_1} + \frac{SD^2{}_2}{n_2}}}$$

Where:

- M_1 and M_2: Sample means for the two groups.
- SD^2_1 and SD^2_2: Variances (squared standard deviations) for the two groups.
- n_1 and n_2: Sample sizes for the two groups.

Key Points

- **Variances:** The formula uses the squared standard deviations (variances) to reflect variability within each group. These variances are added together inside the square root symbol.
- **Sample Sizes:** The formula uses n_1 and n_2 to calculate the contribution of each group's variance relative to its size.
- **Degrees of Freedom:** For the independent-samples *t*-test, the degrees of freedom are calculated as $N - 2$, where N is the total number of participants across both groups.

By combining the differences between the two group means, their variances, and sample sizes, the independent-samples *t*-test determines whether the observed difference is statistically significant.

Example: Independent-Samples *t*-Test

A health psychologist aims to investigate whether individuals who regularly consume junk food differ in their calorie estimates compared to those who rarely eat junk food. Since the psychologist does not have a clear hypothesis about the direction of the difference, a two-tailed test is chosen for the analysis.

The psychologist collects calorie estimation data from two independent groups:

- Junk food eaters (8 participants): 180, 220, 150, 85, 200, 170, 150, 190
  - Mean (M_1) = 168.12
  - Standard deviation (SD_1) = 42.66
- Non-junk food eaters (7 participants): 200, 240, 190, 175, 200, 300, 240
  - Mean (M_2) = 220.71
  - Standard deviation (SD_2) = 41.23

The goal is to determine if there is a statistically significant difference between the two groups' calorie estimates.

Hypotheses

Null Hypothesis (H_0): There is no difference between the population means of the two groups ($\mu_1 = \mu_2$).

Alternative Hypothesis (H_A): There is a difference between the population means of the two groups ($\mu_1 \neq \mu_2$).

The t-statistic formula for independent samples is:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD^2{}_1}{n_1} + \frac{SD^2{}_2}{n_2}}}$$

Where:

- M_1 and M_2 are the sample means of the two groups.
- $SD^2{}_1$ and $SD^2{}_2$ are the variances (squared standard deviations) of the two groups.
- n_1 and n_2 are the sample sizes of the two groups.

Substituting the values into the formula:

$$t = \frac{220.71 - 168.12}{\sqrt{\frac{42.66^2}{8} + \frac{41.23^2}{7}}} = 2.42$$

The calculated t-statistic is $t = -2.42$.

The psychologist inputs the data into statistical software such as SPSS, Excel, or an online t-test calculator. The software calculates a two-tailed p-value of 0.015 for $t = -2.42$ with $df = 13$ (degrees of freedom: $N - 2 = 15 - 2$).

- **Decision:** Since $p = 0.015$ is less than the significance threshold ($\alpha = 0.05$), the psychologist rejects the null hypothesis.
- **Conclusion:** The results suggest that people who eat junk food regularly estimate fewer calories compared to those who rarely eat junk food.

To confirm the results, the psychologist refers to a t-distribution table. For a two-tailed test with $df = 13$ and $\alpha = 0.05$, the critical t value is ± 2.160.

- **Comparison:** The observed t value (-2.42) is more extreme than the critical value ($-2.42 < -2.160$).
- **Decision:** The null hypothesis is rejected.

The results can be reported as follows:

- $t(13) = -2.42, p = .015$

This format includes the degrees of freedom ($df = 13$), the t-statistic ($t = -2.42$), and the p-value ($p = .015$).

The analysis indicates a statistically significant difference between the calorie estimates of junk food eaters and non-junk food eaters. Specifically, junk food eaters provide lower calorie estimates compared to those who rarely eat junk food. This finding could suggest that regular junk food consumption might influence perceptions of calorie content, potentially contributing to dietary misjudgements.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

11.3. THE ANALYSIS OF VARIANCE

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

While *t*-tests are ideal for comparing two means, such as between a sample mean and a population mean, or between the means of two groups or conditions, they are not suitable when more than two groups or conditions need to be compared. For these cases, the analysis of variance (ANOVA) is the most common statistical method.

This section focuses primarily on the one-way ANOVA, which is used in between-subjects designs with a single independent variable. Additionally, it provides a brief overview of other types of ANOVA, such as those used in within-subjects designs and factorial designs.

One-Way ANOVA

The one-way ANOVA is a statistical test used to compare the means of more than two groups (e.g., M_1, M_2...M_G) in a between-subjects design. It is particularly useful when the research involves multiple conditions or groups.

The null hypothesis for a one-way ANOVA is that all group means are equal in the population ($\mu_1 = \mu_2 = \cdots = \mu_G$). The alternative hypothesis asserts that not all means are equal.

The *F* Statistic in ANOVA

The test statistic for ANOVA is called *F*, which is calculated as the ratio of two estimates of population variance derived from the sample data:

1. **Mean Squares Between Groups (MS_B):** This measures variability between the group means.
2. **Mean Squares Within Groups (MS_W):** This measures variability within each group.

The formula for the *F* statistic is:

$F = MS_B / MS_W$

This statistic is useful because we know how it behaves under the null hypothesis. As shown in Figure 11.3.1, the *F*-distribution is unimodal and positively skewed, with most values clustering around 1 when the null hypothesis is true. The shape of the distribution depends on two types of degrees of freedom:

- Between-groups degrees of freedom: $df_B = G - 1$, where G is the number of groups.
- Within-groups degrees of freedom: $df_W = N - G$, where N is the total sample size.

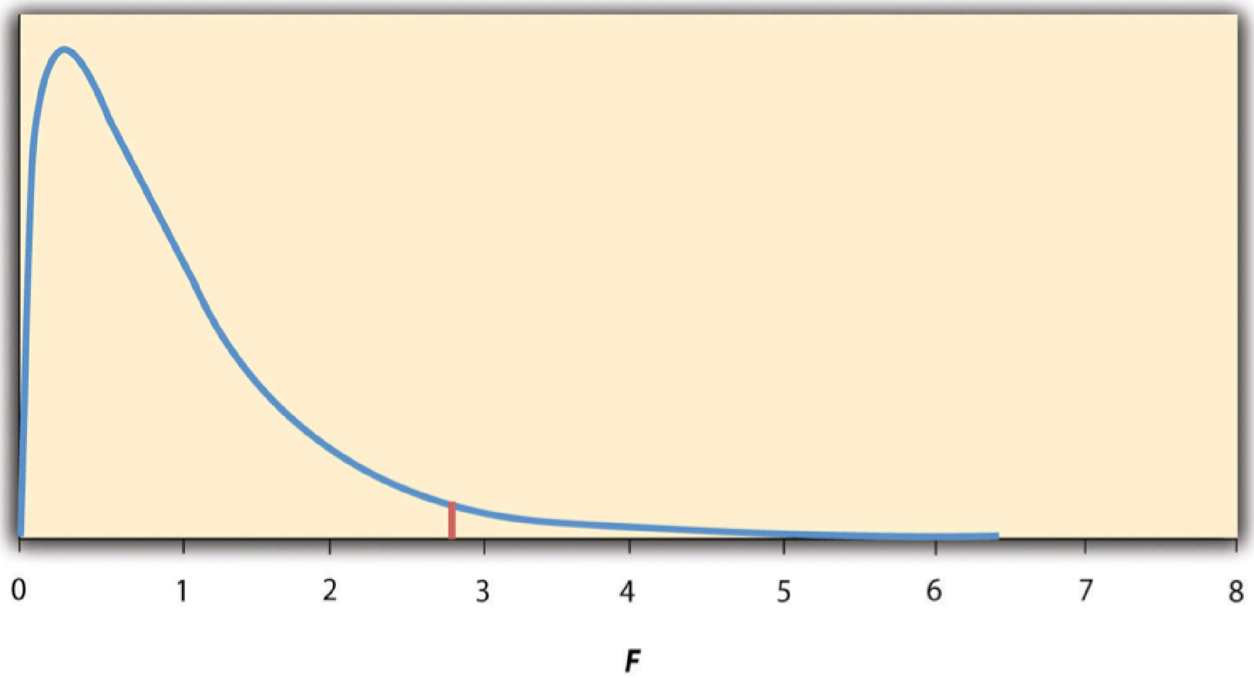


Figure 11.3.1. Distribution of the *F* ratio with 2 and 37 degrees of freedom when the null hypothesis is true. The red vertical line represents the critical value when α is .05, by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Interpreting *F* and the *p*-Value

By knowing the distribution of *F* under the null hypothesis, we can calculate the *p*-value. Statistical software like SPSS or Excel can compute both the *F* statistic and the *p*-value.

- If $p \leq 0.05$, reject the null hypothesis and conclude that there are differences among the group means.
- If $p > 0.05$, retain the null hypothesis, as there is not enough evidence to conclude differences exist.

Using Critical Values to Interpret *F*

In cases where *F* is calculated manually, a table of critical values (like Table 11.3.1) can be used. For example:

- If the computed *F* ratio exceeds the critical value for the given degrees of freedom (df_B and df_W), the null hypothesis is rejected.

- If the computed F ratio is less than the critical value, the null hypothesis is retained.
- Critical values vary based on the degrees of freedom and the significance level ($\alpha = 0.05$).

Table 11.3.1. Table of critical values of *F* when α = .05

			df_B
df_W	2	3	4
8	4.459	4.066	3.838
9	4.256	3.863	3.633
10	4.103	3.708	3.478
11	3.982	3.587	3.357
12	3.885	3.490	3.259
13	3.806	3.411	3.179
14	3.739	3.344	3.112
15	3.682	3.287	3.056
16	3.634	3.239	3.007
17	3.592	3.197	2.965
18	3.555	3.160	2.928
19	3.522	3.127	2.895
20	3.493	3.098	2.866
21	3.467	3.072	2.840
22	3.443	3.049	2.817
23	3.422	3.028	2.796
24	3.403	3.009	2.776
25	3.385	2.991	2.759
30	3.316	2.922	2.690
35	3.267	2.874	2.641
40	3.232	2.839	2.606
45	3.204	2.812	2.579
50	3.183	2.790	2.557
55	3.165	2.773	2.540
60	3.150	2.758	2.525
65	3.138	2.746	2.513
70	3.128	2.736	2.503
75	3.119	2.727	2.494
80	3.111	2.719	2.486
85	3.104	2.712	2.479

90	3.098	2.706	2.473
95	3.092	2.700	2.467
100	3.087	2.696	2.463

Example: One-Way ANOVA

A health psychologist is investigating whether different groups estimate the calorie content of a chocolate chip cookie differently. The three groups in the study are psychology majors, nutrition majors, and professional dietitians. The psychologist collects calorie estimation data from participants in each group and compares the group means using a one-way ANOVA. This statistical test is designed to determine if there are significant differences among the group means in the population.

The calorie estimates from each group are as follows:

- Psychology majors (8 participants): 200, 180, 220, 160, 150, 200, 190, 200
  - Mean (M_1) = 187.50
  - Standard Deviation (SD_1) = 23.14
- Nutrition majors (9 participants): 190, 220, 200, 230, 160, 150, 200, 210, 195
  - Mean (M_2) = 195.00
  - Standard Deviation (SD_2) = 27.77
- Dieticians (8 participants): 220, 250, 240, 275, 250, 230, 200, 240
  - Mean (M_3) = 238.13
  - Standard Deviation (SD_3) = 22.35

From the descriptive statistics, it appears that dieticians provide significantly higher calorie estimates than both psychology and nutrition majors.

Hypotheses

- **Null Hypothesis (H_0):** The mean calorie estimates for the three groups are equal in the population ($\mu_1 = \mu_2 = \mu_3$).
- **Alternative Hypothesis (H_A):** At least one of the group means is significantly different from the others.

The psychologist decides to run a one-way ANOVA to statistically test whether the group means are significantly different in the population. The psychologist uses statistical software (e.g., SPSS, Excel, or another tool) to conduct the one-way ANOVA. The software calculates the F ratio and the p-value to determine whether the differences among the group means are statistically significant.

The results of the ANOVA are summarised in the following table (Table 11.3.2):

Table 11.3.2. ANOVA results

Source of variation	***SS***	***df***	***MS***	***F***	***p-value***	$F_{critical}$
Between groups	11,943.75	2	5,971.875	9.916234	0.000928	3.4668
Within groups	12,646.88	21	602.2321			
Total	24,590.63	23				

Key results from the table:

- **MS_B (mean squares between groups):** 5,971.88
- **MS_W (mean squares within groups):** 602.23
- ***F* (*F* ratio):** 9.92
- ***p*-value:** 0.0009

Interpretation of Results

- **Decision rule:** The critical value of F ($F_{critical}$) at $\alpha = 0.05$, with 2 and 21 degrees of freedom, is 3.467. The computed *F* ratio (9.92) exceeds this critical value, indicating that the result is statistically significant.
- ***p*-value:** The *p*-value (0.0009) is much smaller than 0.05, providing strong evidence to reject the null hypothesis.
- **Conclusion:** The psychologist concludes that the mean calorie estimates for the three groups are not equal in the population. This suggests significant differences in how the three groups estimate calorie content.

The results provide strong evidence that dieticians estimate calories differently from the other two groups. However, the ANOVA does not specify which group means are different from each other. To identify the specific group differences, post hoc comparisons (e.g., Tukey's HSD test) would need to be conducted.

The results can be reported as follows:

- $F(2, 21) = 9.92, p = 0.0009$

This format indicates:

- Degrees of freedom ($df_B = 2$, $df_W = 21$)
- *F*-statistic ($F = 9.92$)
- *p*-value ($p = 0.0009$)

The sum of squares (SS) values for "Between Groups" and "Within Groups" are intermediate calculations

used to find the MS_B and MS_W values but are not typically reported in the final results.
If the *F* ratio had been computed manually, the psychologist could use a table of critical values (like Table 11.3.1) to confirm the statistical significance.

Post Hoc Comparisons: Follow-Up Tests After ANOVA

When the null hypothesis is rejected in a one-way ANOVA, we conclude that the group means are not all equal in the population. However, this result does not clarify the specific nature of the differences. For instance, in a study with three groups:

1. All three means might differ significantly from one another. For example, the mean calorie estimates for psychology majors, nutrition majors, and dieticians could all be distinct.
2. Only one mean might differ significantly. For example, dieticians' calorie estimates might differ significantly from those of psychology and nutrition majors, while the psychology and nutrition majors' means remain similar.

Because ANOVA does not specify which groups differ, significant results are typically followed by post hoc comparisons to identify specific group differences.

Why Not Just Use Multiple *t*-Tests?

A straightforward approach might be to run a series of independent-samples *t*-tests, comparing each group's mean to every other group. However, this method has a critical flaw: increased risk of type I errors. Each *t*-test has a 5% chance of incorrectly rejecting the null hypothesis when it is true. Conducting multiple *t*-tests increases the cumulative probability of making at least one error. For example, with three groups, conducting three *t*-tests raises the likelihood of a mistake beyond the acceptable 5% level.

To control for this inflated error risk, researchers use modified *t*-test procedures designed to keep the overall error rate within acceptable limits (close to 5%). These methods include:

- **Bonferroni Procedure:** Adjusts the significance threshold for each test based on the number of comparisons.
- **Fisher's Least Significant Difference (LSD) Test:** Performs comparisons but with adjustments to maintain accuracy.
- **Tukey's Honestly Significant Difference (HSD) Test:** Specifically designed for pairwise comparisons to control error rates.

While the technical details of these methods are beyond the scope of this explanation, their purpose is straightforward: to minimise the chance of mistakenly rejecting a true null hypothesis while identifying meaningful group differences.

Repeated-Measures ANOVA

The one-way ANOVA is designed for between-subjects designs, where the means being compared come from separate groups of participants. However, it is not suitable for within-subjects designs, where the same participants are tested under different conditions or at different times. In these cases, a repeated-measures ANOVA is used instead.

The fundamental principles of repeated-measures ANOVA are similar to those of the one-way ANOVA. The primary distinction lies in how the analysis accounts for variability within the data:

Handling Individual Differences

Measuring the dependent variable multiple times for each participant allows the analysis to account for stable individual differences. These are characteristics that consistently vary among participants but remain unchanged across conditions. For example:

- In a reaction-time study, some participants might be naturally faster or slower due to stable factors like their nervous system or muscle response.
- In a between-subjects design, these differences add to the variability within groups, increasing the mean squares within groups (MS_W). A higher MS_W results in a lower F value, making it harder to detect significant differences.
- In a repeated-measures design, these stable individual differences are measured and subtracted from MS_W, reducing its value.

Increased Sensitivity

By reducing the value of MS_W, repeated-measures ANOVA produces a higher F value, making the test more sensitive to detecting actual differences between conditions.

This approach leverages the fact that participants serve as their own control, allowing the analysis to isolate and remove irrelevant variability caused by stable individual differences. As a result, repeated-measures ANOVA offers a more precise and powerful way to detect meaningful effects in within-subjects designs.

Factorial ANOVA

When a study includes more than one independent variable, the appropriate statistical approach is the factorial ANOVA. This method builds on the principles of the one-way and repeated-measures ANOVAs but is designed to handle the complexities of factorial designs.

Multiple Effects Analysis

Factorial ANOVA examines not only the impact of each independent variable (main effects) but also how these variables interact (interaction effects). For each:

- It calculates an F ratio to measure the variability explained by the effect.
- It provides a p value to indicate whether the effect is statistically significant.

Main Effects

These represent the individual influence of each independent variable on the dependent variable. For example:

- In a calorie estimation study, one main effect might measure whether a participant's major (psychology vs. nutrition) influences calorie estimates.
- Another main effect might examine whether the type of food (cookie vs. hamburger) affects the estimates.

Interaction Effects

The interaction effect explores whether the influence of one independent variable depends on the level of the other. For instance:

- Does the participant's major (psychology vs. nutrition) affect calorie estimates differently depending on the type of food (cookie vs. hamburger)?

Customising for Study Design

The factorial ANOVA is adaptable and must be tailored based on the study design:

- **Between-subjects design:** Participants are assigned to separate groups based on the levels of the independent variables.
- **Within-subjects design:** The same participants experience all levels of the independent variables.
- **Mixed design:** Combines between-subjects and within-subjects elements.

Imagine a health psychologist investigates calorie estimation using two independent variables: participant major (psychology vs. nutrition) and food type (cookie vs. hamburger). A factorial ANOVA would calculate:

- An F ratio and p value for the main effect of major.
- An F ratio and p value for the main effect of food type.
- An F ratio and p value for the interaction between major and food type.

This analysis provides a detailed understanding of how the independent variables individually and collectively influence calorie estimates, making the factorial ANOVA a powerful tool for complex research designs.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

11.4. CORRELATION COEFFICIENTS

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

When studying relationships between quantitative variables, Pearson's *r* (the correlation coefficient) is a widely used measure that quantifies the strength and direction of the relationship between two variables. However, to determine whether this observed relationship is statistically significant, that is, unlikely to have occurred by chance. A formal test of the correlation coefficient is required. This test follows the same fundamental principles as other null hypothesis tests, aiming to make inferences about the population based on the sample data.

The null and alternative hypotheses for testing the significance of Pearson's r are as follows:

- **Null hypothesis ($\rho = 0$):** There is no relationship between the variables in the population. The population correlation, represented by the Greek letter rho (ρ), is equal to zero.
- **Alternative hypothesis ($\rho \neq 0$):** There is a relationship between the variables in the population. The population correlation is not equal to zero.

The null hypothesis assumes that any observed correlation in the sample is due to random variation, while the alternative hypothesis asserts that the correlation reflects a true relationship in the population.

Statistical Approach to Testing Pearson's *r*

There are two primary methods for conducting the test:

1. **Using a *t* score:** The sample correlation coefficient (*r*) can be converted into a *t* score using the formula for a *t* statistic. This *t* score is evaluated against a *t* distribution with $N - 2$ degrees of freedom, where *N* is the sample size. From this point, the procedure aligns with the standard approach for a *t*-test. The *t* score provides a basis for calculating the *p* value, which helps determine whether the null hypothesis should be rejected.
2. **Direct Use of Pearson's *r*:** Because of the way it is computed, Pearson's *r* itself can function as the test statistic. Modern statistical tools like Excel, SPSS, or online calculators typically compute Pearson's *r* and provide the corresponding *p* value automatically. These tools simplify the process,

allowing researchers to quickly assess the significance of the correlation without manual calculations.

Interpreting the Results

If the *p* value is ≤ 0.05, we reject the null hypothesis and conclude there is a relationship between the variables in the population. However, if the *p* value is > 0.05, we retain the null hypothesis, meaning there is not enough evidence to conclude a relationship exists.

If calculating by hand, you can compare the sample's correlation coefficient to critical values in a table like Table 11.4.1. The critical value depends on the sample size (*N*) and whether the test is one-tailed or two-tailed. If the sample's correlation coefficient is more extreme than the critical value, it is statistically significant.

Table 11.4.1. Table of critical values of Pearson's *r* when α = .05

	Critical value of *r*	
N	One-tailed	Two-tailed
5	.805	.878
10	.549	.632
15	.441	.514
20	.378	.444
25	.337	.396
30	.306	.361
35	.283	.334
40	.264	.312
45	.248	.294
50	.235	.279
55	.224	.266
60	.214	.254
65	.206	.244
70	.198	.235
75	.191	.227
80	.185	.220
85	.180	.213
90	.174	.207
95	.170	.202
100	.165	.197

Example: Testing a Correlation Coefficient

A health psychologist is exploring whether there is a correlation between individuals' calorie estimates for food and their body weight. She does not have a specific expectation about whether the relationship will be positive or negative. Therefore, she opts for a two-tailed test, which allows her to detect a correlation in either direction.

The psychologist collects data from a sample of 22 university students, recording each student's calorie estimates for a particular food item along with their body weight. Using this data, she calculates the correlation coefficient, Pearson's *r*, which measures the strength and direction of the relationship. The resulting value is −0.21, indicating a weak negative correlation. This suggests that, in this sample, higher calorie estimates tend to be associated with lower body weights, though the relationship is not strong.

To assess whether this observed correlation is statistically significant, meaning that it likely reflects a true relationship in the population rather than random chance, the psychologist uses statistical software. The software computes the *p* value associated with her correlation coefficient and reports it as 0.348.

- **Interpreting the *p* value:** The *p* value represents the probability of obtaining a sample correlation as extreme as −0.21, or more extreme, if the null hypothesis is true (i.e., if there is no correlation in the population). In this case, the *p* value is 0.348, which is substantially greater than the conventional significance threshold of 0.05.

Because the *p* value exceeds 0.05, the psychologist retains the null hypothesis, concluding that the data does not provide sufficient evidence to suggest a significant relationship between calorie estimates and weight in the population.

If the psychologist were calculating the test manually, she would refer to a table of critical values for Pearson's *r* to confirm whether the correlation is statistically significant. For a sample size of 22 participants, the degrees of freedom are calculated as $N - 2$, or 20.

- **Critical value for two-tailed test:** From the critical value table (e.g., Table 11.4.1), she finds that for a two-tailed test with $\alpha = 0.05$ and 20 degrees of freedom, the critical value of *r* is ±0.444.

To determine significance, the absolute value of the sample correlation coefficient ($|-0.21| = 0.21$) is compared to the critical value:

- If $|r| > 0.444$, the correlation is statistically significant.
- If $|r| \leq 0.444$, the correlation is not statistically significant.

Since 0.21 is less extreme than the critical value of ±0.444, the psychologist confirms that the *p* value is greater than 0.05. This further supports her decision to retain the null hypothesis.

Based on the analysis, the psychologist concludes that there is no statistically significant relationship between calorie estimates and weight in the population. The weak negative correlation observed in the sample (−0.21) could be due to random sampling variability rather than a true association.

Chapter Attribution

Content adapted, with editorial changes, from:

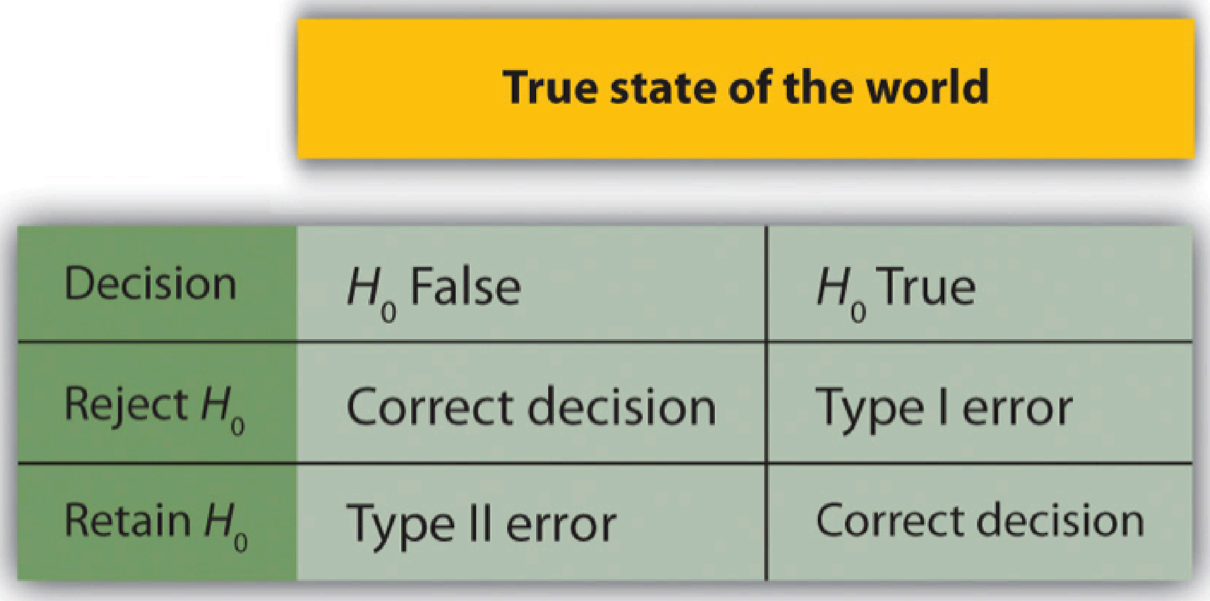
Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

11.5. TYPE I AND TYPE II ERRORS

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Null hypothesis testing is a statistical process used by researchers to draw conclusions about a population based on the analysis of sample data. The ultimate goal is to determine whether there is enough evidence to reject the null hypothesis (which assumes no effect or relationship) or to retain it. However, these conclusions are not always guaranteed to be correct because of the inherent limitations of sampling and statistical testing. Errors can occur, leading researchers to make incorrect decisions about the null hypothesis.

To understand these potential errors, consider the four possible outcomes of a null hypothesis test, as illustrated in Figure 11.5.1. These outcomes arise from the interplay between the actual state of the null hypothesis in the population (true or false) and the researcher's decision to reject or retain the null hypothesis based on sample data.



	True state of the world	
Decision	H_0 False	H_0 True
Reject H_0	Correct decision	Type I error
Retain H_0	Type II error	Correct decision

Figure 11.5.1. Two types of correct decisions and two types of errors in null hypothesis testing by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

Four Possible Outcomes in Null Hypothesis Testing

Correct Decision: Rejecting a False Null Hypothesis

- **Scenario:** The null hypothesis is false in the population, meaning there is an actual effect or relationship. The researcher correctly rejects the null hypothesis based on the sample data.
- **Significance:** This is the ideal outcome, indicating that the test successfully identified a real effect or relationship in the population.
- **Example:** A study finds a significant difference in weight loss between two diets, and this difference also exists in the population.

Correct Decision: Retaining a True Null Hypothesis

- **Scenario:** The null hypothesis is true in the population, meaning there is no effect or relationship. The researcher correctly retains the null hypothesis, concluding that the sample data do not provide sufficient evidence of an effect.
- **Significance:** This outcome reflects the proper use of statistical analysis, avoiding overinterpretation of random sampling variability.
- **Example:** A study finds no significant difference in test scores between two teaching methods, and there is no actual difference in the population.

Type I Error: Rejecting a True Null Hypothesis

- **Scenario:** The null hypothesis is true in the population, but the researcher erroneously rejects it based on the sample data.
- **Definition:** This is known as a Type I error, where a false positive occurs, and the test identifies an effect or relationship that does not actually exist.
- **Cause:** Type I errors are typically caused by random sampling variability and are influenced by the significance level (α) set by the researcher. For example, with $\alpha = 0.05$, there is a 5% chance of committing a Type I error.
- **Example:** A study concludes that a new medication improves patient outcomes, but in reality, the medication has no effect.

Type II Error: Retaining a False Null Hypothesis

- **Scenario:** The null hypothesis is false in the population, but the researcher fails to reject it based on the sample data.
- **Definition:** This is known as a Type II error, where a false negative occurs, and the test fails to detect an effect or relationship that actually exists.
- **Cause:** Type II errors are often due to inadequate statistical power, which may result from a small

sample size, a weak effect, or high variability in the data.

- **Example:** A study finds no significant improvement in learning outcomes using a new teaching method, even though it is effective in the population.

Type I and Type II Errors

Errors in statistical hypothesis testing are an inherent part of the research process, primarily because conclusions about populations are drawn from sample data. These errors, referred to as Type I and Type II errors, arise due to sampling variability, limitations in research design, or decisions about statistical thresholds.

Type I Error: False Positive

A Type I error occurs when a researcher rejects the null hypothesis (H_0) even though it is true in the population. This type of error leads to a false positive result, suggesting the presence of an effect or relationship that does not actually exist.

- **Cause:** Even when there is no true relationship in the population, random sampling error can occasionally produce extreme sample results. These extreme results can lead to the rejection of the null hypothesis.
- **Role of Significance Level (α):** The significance level, α, determines the threshold for rejecting the null hypothesis. It represents the probability of committing a Type I error. For instance, if $\alpha = 0.05$, there is a 5% chance of rejecting a true null hypothesis purely due to random sampling error.
    - This is why α is often called the "Type I error rate".
- **Implications:** A Type I error can lead researchers to falsely conclude that an intervention, treatment, or relationship is effective when it is not. This can have significant consequences, such as implementing ineffective medical treatments or drawing incorrect scientific conclusions.
- **Example:** Suppose a clinical trial tests a new medication and finds a statistically significant improvement in patients' outcomes. If this result is a Type I error, the medication does not actually provide any benefit, and the observed effect is due to random chance.

Type II Error: False Negative

A Type II error occurs when a researcher fails to reject the null hypothesis (H_0) even though it is false in the population. This type of error results in a false negative, meaning the researcher overlooks a real effect or relationship.

- **Cause:** Type II errors often stem from a lack of statistical power in the research design. Statistical power is the probability of correctly rejecting the null hypothesis when it is false.

- **Factors reducing power:**
  - Small sample sizes
  - High variability in the data
  - Weak effects or relationships in the population
- **Balancing Type II Errors and α:** Adjusting the significance level can influence the likelihood of a Type II error. Lowering α (e.g., from 0.05 to 0.01) makes it harder to reject the null hypothesis, thereby increasing the risk of Type II errors. Conversely, raising α (e.g., to 0.10) makes it easier to reject the null hypothesis, reducing the likelihood of Type II errors but increasing the risk of Type I errors.
- **Implications:** A Type II error can prevent researchers from identifying effective treatments, interventions, or relationships. This oversight can lead to missed opportunities for advancements or improvements in various fields.
- **Example:** A study investigates whether a new teaching method improves student performance. If the study fails to detect a significant difference due to a Type II error, the method might be incorrectly dismissed as ineffective, even though it could benefit students.

Balancing Type I and Type II Errors

Researchers must carefully balance the risks of Type I and Type II errors when designing studies and interpreting results. The commonly used significance level of $\alpha = 0.05$ reflects a compromise between these risks, aiming to keep both error rates at acceptable levels. However, the appropriate balance may vary depending on the context of the research.

- **Lowering Type I Errors:** Reducing α (e.g., to 0.01) decreases the likelihood of false positives but increases the probability of Type II errors. This approach is often used in fields where the consequences of a Type I error are severe, such as drug approvals or medical diagnostics.
- **Reducing Type II Errors:** Increasing statistical power through larger sample sizes, better-controlled experiments, or stronger manipulations can reduce the likelihood of Type II errors. This ensures a greater chance of detecting real effects or relationships when they exist.

Implications of Errors in Research and Addressing Bias in the Literature

Errors in null hypothesis testing, such as Type I and Type II errors, have significant implications for interpreting research findings and the advancement of scientific knowledge. Researchers must remain aware of these potential errors and their consequences to ensure their conclusions are reliable and accurate. Below, we explore these implications in greater detail and discuss solutions to mitigate their impact.

Caution in Interpreting Results

Every study, no matter how well-designed, carries the risk of a Type I or Type II error (Figure 11.5.2). This inherent uncertainty means researchers should approach their findings with caution, recognising that an observed result could either falsely indicate a relationship (Type I error) or fail to detect a real one (Type II error).

Importance of Replication:

- Replication is a cornerstone of reliable science. Each successful replication, where independent researchers conduct the same study and obtain similar results, bolsters confidence that a finding reflects a genuine phenomenon rather than an error.
- For example, if a study finds a significant relationship between exercise and improved memory performance, replication ensures the result is not due to random chance, sampling variability, or methodological flaws.
- Meta-analyses, which aggregate findings from multiple studies, further enhance reliability by reducing the influence of outliers or single-study errors.

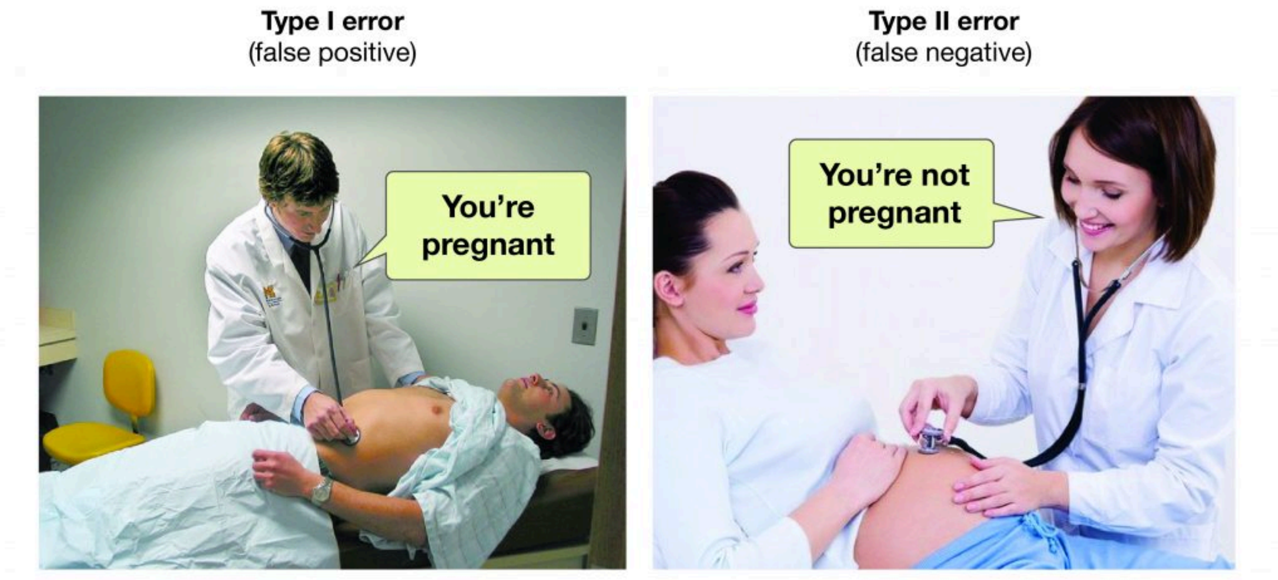


Figure 11.5.2. A humorous example of how Type I and Type II errors could play out in pregnancy exams by R. S. Jhangiani et al. in Research Methods in Psychology 4e is used under a CC BY-NC-SA licence

The File Drawer Problem

One significant issue in scientific research is the file drawer problem, first articulated by Rosenthal in 1979. This problem arises from the tendency for researchers and journals to prioritise publishing statistically significant results while relegating non-significant findings to obscurity. In other words, placing them in a "file drawer" figuratively.

- **Inflated Proportion of Type I Errors in Published Research:** When only significant results are published, the research literature becomes biased, overrepresenting false positives (Type I errors). Studies that find no relationship or fail to reach statistical significance, even when accurately reflecting the population, are underreported.
- **Impact on True Relationships:** This bias can also distort the perceived strength of relationships. For instance, if a weak positive correlation exists between two variables in the population, published studies may disproportionately report stronger correlations due to random sampling error and the file drawer effect. This creates a misleading picture, exaggerating the strength of the true relationship.
- **Example:** Imagine multiple studies examining the relationship between caffeine intake and cognitive performance. Some studies show weak or no effect, while a few, due to sampling variability, report moderate or strong effects. If only the latter are published, readers might wrongly conclude that caffeine has a robust impact on cognition, while the true effect is minimal or non-existent.

Addressing the File Drawer Problem

The file drawer problem is deeply embedded in traditional research and publishing practices. However, researchers and journals are increasingly adopting innovative solutions to counteract this bias and ensure a more balanced representation of findings.

Registered Reports: Shifting the Focus to Research Quality

Registered reports are a publishing format where journals evaluate a study based on its research question and methodology, rather than its results. Researchers submit their study design and hypotheses for review before data collection. If the design is deemed rigorous, the journal commits to publishing the results regardless of whether they are significant or non-significant.

- Benefits:
  - Reduces the bias toward publishing significant results.
  - Encourages transparent and well-planned research.
  - Ensures that non-significant findings are valued equally, as they contribute to the broader understanding of a topic.
- Example:
  - The Centre for Open Science promotes registered reports through initiatives like this platform, fostering fair evaluation across scientific disciplines.

Sharing Non-Significant Results

Researchers can actively address publication bias by sharing non-significant findings through alternative avenues:

- **Public Repositories:** Platforms like the Open Science Framework allow researchers to upload and share raw data, methodologies, and results, making them accessible to the scientific community.
- **Professional Conferences:** Conferences provide opportunities for researchers to present non-significant findings and foster discussions about their implications.
- **Dedicated Journals:** Some journals specifically focus on non-significant findings, such as the Journal of Articles in Support of the Null Hypothesis. These outlets help balance the literature by highlighting studies that may otherwise remain unpublished.

Promoting Cultural Change in Science

Addressing the file drawer problem requires a cultural shift in how research is evaluated and rewarded. By recognising the value of all results, whether significant or not, scientific disciplines can move toward a more complete and accurate understanding of phenomena.

p-Hacking and Its Consequences

In 2014, researchers Uri Simonsohn, Leif Nelson, and Joseph Simmons brought significant attention to a widespread but problematic practice in psychological research known as *p*-hacking. This practice involves researchers making intentional or unintentional adjustments to their data analyses in order to achieve statistically significant *p*-values, which are often considered necessary for publication. While *p*-hacking may seem harmless on the surface, it has significant consequences for the reliability and validity of scientific findings.

What is *p*-Hacking?

p-hacking encompasses a range of questionable research practices aimed at "hacking" the *p*-value, which represents the probability of observing a result as extreme as the one obtained, assuming the null hypothesis is true. A *p*-value of 0.05 or less is often treated as the threshold for statistical significance, making it a key benchmark for whether results are published. Researchers engaging in *p*-hacking manipulate their analyses to cross this threshold, even if the underlying data do not genuinely support such conclusions.

Examples of *p*-Hacking Tactics:

1. **Selective Removal of Outliers:** Researchers may exclude certain data points that seem to weaken

the statistical significance of their findings, labelling them as “outliers”.

2. **Data Snooping:** Running multiple statistical tests on the same dataset and selectively reporting only the tests that yield significant results.
3. **Selective Reporting of Variables:** Including or excluding variables from the analysis until significant results are obtained.
4. **Stopping Data Collection Early or Extending It:** Stopping data collection as soon as a significant *p*-value is achieved or continuing to collect data until significance is reached.
5. **Post Hoc Hypothesis Formulation:** Formulating hypotheses after the results are known, presenting them as if they were predicted in advance.

Consequences of *p*-Hacking

The widespread use of *p*-hacking undermines the foundation of scientific research by increasing the likelihood of Type I errors, or false positives, where researchers mistakenly conclude that a relationship or effect exists when it does not. This leads to several cascading problems:

1. **Inflated False Positives in the Literature:** *p*-hacking inflates the number of published studies that report significant findings, even when these findings are due to random chance rather than real effects. This distorts the scientific literature, making it difficult for other researchers to discern true relationships or effects.
2. **Reduced Replicability:** Studies influenced by *p*-hacking are less likely to replicate. When other researchers attempt to reproduce these findings, they often fail because the original results were artifacts of manipulated analyses rather than genuine phenomena.
3. **Erosion of Public and Academic Trust:** When *p*-hacking is revealed, it undermines trust in science as a whole. Both the public and the academic community may become sceptical of research findings, questioning whether they are genuine or the result of data manipulation.
4. **Misdirected Resources and Efforts:** Researchers, funding agencies, and practitioners may waste time and resources pursuing ideas or interventions based on spurious findings. This slows progress and diverts attention from potentially meaningful discoveries.

The 2014 Exposé and Its Impact

The work of Simonsohn, Nelson, and Simmons served as a wake-up call for the scientific community. Their research demonstrated how easily researchers could manipulate their analyses to produce significant *p*-values, even from entirely random datasets. Their findings highlighted the systemic vulnerability of research practices in psychology and other fields, sparking an ongoing conversation about the credibility of scientific research.

However, *p*-hacking is not always deliberate. In many cases, researchers may unknowingly engage in these practices due to pressure to publish or misunderstandings about proper statistical procedures. The emphasis on statistical significance as a criterion for publication incentivises *p*-hacking, creating a culture where "getting the right result" becomes more important than accurate reporting.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

11.6. STATISTICAL POWER

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Statistical power is a critical concept in research methodology, reflecting the likelihood of correctly rejecting the null hypothesis when it is false. In other words, it measures a study's ability to detect a true effect or relationship in the population based on the sample data. Several factors influence statistical power, including the sample size, the expected strength of the relationship, and the chosen significance level (α).

Consider a scenario where a researcher conducts a study with 50 participants to investigate a population correlation, expecting Pearson's *r* to be +0.30. The statistical power for this study is 0.59. This indicates a 59% probability of correctly rejecting the null hypothesis if the population correlation is indeed +0.30. The probability of failing to reject a false null hypothesis, a Type II error, is the complement of power. In this example, the likelihood of making a Type II error is calculated as 1 − 0.59 = 0.41, or 41%.

Statistical power serves as an essential safeguard against Type II errors, where true effects in the population go undetected. To minimise such errors, researchers aim for an adequate power level, typically set at 0.80. This standard implies an 80% chance of detecting a true effect, assuming it exists, which balances reliability and resource efficiency.

How to Calculate Statistical Power

While calculating statistical power involves complex formulas tailored to specific research designs, modern tools simplify the process. Researchers can use statistical software or online calculators by providing inputs like:

- **Sample size (*N*):** The number of participants or observations.
- **Effect size:** The expected strength of the relationship (e.g., Cohen's *d* or Pearson's *r*).
- **Significance level (α):** Typically set at 0.05, reflecting a 5% risk of a Type I error.

Table 11.6.1 illustrates the sample sizes required to achieve a power of 0.80 for various effect sizes in two-tailed independent-samples *t*-tests and tests of Pearson's *r*:

Table 11.6.1. Sample sizes needed to achieve statistical power of .80 for different expected relationship strengths for an independent-samples *t* test and a test of Pearson's *r*

Null Hypothesis Test		
Relationship Strength	Independent-Samples *t*-Test	Test of Pearson's *r*
Strong (*d* = .80, *r* = .50)	52	28
Medium (*d* = .50, *r* = .30)	128	84
Weak (*d* = .20, *r* = .10)	788	782

These numbers highlight a critical point: detecting weak relationships requires much larger sample sizes compared to strong or medium relationships.

What If Power Is Inadequate?

Insufficient statistical power undermines the reliability of research findings, increasing the risk of false negatives (Type II errors). For instance, imagine a researcher conducting a between-subjects experiment with 20 participants in each group and expecting a medium effect size (*d* = 0.50). The statistical power of this study is only 0.34. This means there is just a 34% chance of detecting a true effect, leaving a 66% chance of missing it. Such low power renders the study unreliable, risking wasted time, resources, and misleading conclusions.

Strategies for Increasing Statistical Power

Improving statistical power is crucial for ensuring robust and reliable findings. Two primary strategies can help:

Increase the Strength of the Relationship

Researchers can enhance the effect size by:

- **Strengthening manipulations:** For example, using more intense experimental conditions.
- **Controlling extraneous variables:** Reducing variability caused by noise or irrelevant factors improves the clarity of the observed effect.
- **Switching designs:** Using a within-subjects design instead of a between-subjects design can reduce error variance, increasing power.

Increase the Sample Size

The most common and straightforward approach to boosting power is collecting more data. A larger sample reduces the influence of random error, making it easier to detect true relationships. Importantly, for any given effect size, there is always a sample size large enough to achieve adequate power.

Tools for Computing Statistical Power

To aid researchers in planning studies with sufficient power, various tools are available for free or online:

- **Russ Lenth's Power and Sample Size Page**. A user-friendly online tool where researchers can calculate power or determine the required sample size based on their study parameters.
- **G*Power**. A comprehensive, free software program offering advanced capabilities for computing power, effect sizes, and sample size requirements for a wide range of statistical tests.

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

11.7. CRITICISMS OF NULL HYPOTHESIS TESTING

By Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler and Dana C. Leighton, adapted by Marc Chao and Muhamad Alif Bin Ibrahim

Null hypothesis testing is a cornerstone of modern research methodology, yet it has faced increasing scrutiny due to its conceptual and practical limitations. These criticisms range from widespread misunderstandings of key concepts, such as the *p*-value, to fundamental concerns about the method's logic and utility in drawing meaningful conclusions. Despite its shortcomings, null hypothesis testing continues to be defended and widely used, while alternative approaches are gaining traction in research communities.

Misinterpretations of the *p*-Value

One of the most pervasive issues with null hypothesis testing is the misinterpretation of the *p*-value. Many researchers erroneously believe that the *p*-value represents the probability that the null hypothesis is true. In reality, the *p*-value indicates the likelihood of observing a sample result as extreme as the one obtained, assuming the null hypothesis is true. This misunderstanding often leads to overconfidence in research findings and misinformed conclusions. A related misconception involves the belief that $1 - p$ equates to the probability of successfully replicating a significant result. For example, a study by Oakes (1986) revealed that 60% of professional researchers incorrectly thought a *p*-value of 0.01 in an independent-samples *t*-test (with 20 participants per group) implied a 99% chance of replication. This is far from accurate. Statistical power, which is a measure of the likelihood of detecting a true effect, shows that even with a large population effect, replicating a result with 99% probability requires considerably larger sample sizes than typically used. For instance, achieving 80% power for a medium effect size requires 26 participants per group, while 99% power demands 59 participants per group. These figures highlight how reliance on *p*-values can create misleading expectations about replicability.

Criticism of Rigid *p*-Value Thresholds

Another concern is the strict reliance on the $p < 0.05$ threshold to determine statistical significance. This rigid boundary often leads to arbitrary distinctions between "significant" and "non-significant" results. For example, two studies with nearly identical findings, one with $p = 0.04$ and another with $p = 0.06$, might be judged very differently. The former could be viewed as important and publishable, while the latter might be dismissed. This convention not only stifles valuable research but also exacerbates problems like the file drawer issue, where non-significant results remain unpublished.

Limitations of Null Hypothesis Testing

A deeper criticism questions the fundamental logic of null hypothesis testing. Rejecting the null hypothesis merely indicates that there is some nonzero relationship in the population, without specifying the strength or nature of that relationship. This lack of precision is seen as uninformative. Critics argue that, in many cases, the null hypothesis (e.g., $d = 0$ or $r = 0$) is unlikely ever to be strictly true, as any relationship, however minute, will deviate from zero if measured with enough precision. Consequently, rejecting the null hypothesis may reveal little that is not already assumed. For example, this would be akin to a chemist only determining that temperature affects gas volume without providing a detailed equation to describe the relationship.

Defences of Null Hypothesis Testing

Despite these criticisms, null hypothesis testing has defenders. Robert Abelson (1995) argued that when properly understood and executed, it provides a robust framework for research. Particularly in new areas of study, null hypothesis testing offers a systematic way to demonstrate that results are not merely due to chance, lending credibility to new findings.

The End of *p*-Values?

In 2015, the editors of Basic and Applied Social Psychology announced a ban on null hypothesis testing and related statistical procedures (Tramifmow & Marks, 2015). Authors were still permitted to include *p*-values in their submissions, but the editors committed to removing them before publication. While they did not suggest an alternative statistical method to replace null hypothesis testing, they emphasised the importance of relying on descriptive statistics and effect sizes instead.

Although this decision has not been widely adopted in the broader research community, it sparked significant discussion. By challenging the long-standing “gold standard” of statistical validity, the editors invited psychologists to critically reconsider how knowledge is established and communicated within the field. This debate continues to influence conversations about the role and limitations of statistical methods in scientific research.

What Should Be Done?

Even supporters of null hypothesis testing acknowledge its flaws, but what can researchers do to address these issues? The APA Publication Manual offers several recommendations to improve the practice.

One suggestion is to accompany every null hypothesis test with an effect size measure, such as Cohen's *d* or Pearson's *r*. This addition provides an estimate of the strength of the relationship in the population, rather than simply indicating whether a relationship exists. This is important because a *p*-value alone cannot measure relationship strength, as it is influenced by sample size. For instance, even a very weak relationship can appear statistically significant if the sample size is large enough.

Another recommendation is to use confidence intervals instead of null hypothesis tests. A confidence interval represents a range of values that is likely to include the population parameter a certain percentage of the time (usually 95%). For example, if a sample of 20 students has a mean calorie estimate of 200 for a chocolate chip cookie with a 95% confidence interval of 160 to 240, there is a 95% chance that the true population mean lies within this range. Confidence intervals are often easier to interpret than null hypothesis tests and provide useful information for those who still wish to conduct such tests. For instance, in the example above, the sample mean of 200 is statistically significantly different at the .05 level from any hypothetical population mean outside the confidence interval, such as 250.

Finally, more radical solutions propose replacing null hypothesis testing altogether. Bayesian statistics is one such alternative. In this approach, researchers assign initial probabilities to the null hypothesis and alternative hypotheses before conducting a study, then update these probabilities based on the observed data. While Bayesian methods are gaining attention, it is too soon to determine if they will become standard in psychological research. For now, null hypothesis testing, which is enhanced by effect size measures and confidence intervals, remains the predominant method.

References

Abelson, R. P. (1995). *Statistics as principled argument.* Lawrence Erlbaum Associates.

Oakes, M. W. (1986). *Statistical inference: A commentary for the social and behavioural sciences*. Wiley.

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*(1), 1–2. https://doi.org/10.1080/01973533.2015.1012991

Chapter Attribution

Content adapted, with editorial changes, from:

Research methods in psychology, (4th ed.), (2019) by R. S. Jhangiani et al., Kwantlen Polytechnic University, is used under a CC BY-NC-SA licence.

11.8. SUMMARY

By Marc Chao

Summary

Statistical analysis in psychological research bridges sample data and broader population inferences, utilising descriptive measures like means or correlation coefficients to estimate parameters. Sampling variability introduces potential errors, which null hypothesis significance testing (NHST) addresses by determining whether observed patterns reflect true population relationships or random chance. NHST evaluates the null hypothesis (no relationship) against the alternative hypothesis (a genuine relationship), using the *p*-value to guide conclusions. A *p*-value below 0.05 typically indicates statistical significance, but researchers must distinguish this from practical or clinical significance, which gauges real-world relevance. Tools like *t*-tests and analysis of variance (ANOVA) provide foundational methods for comparing means and exploring group differences, while Pearson's *r* quantifies relationships between variables, requiring significance testing to validate findings.

Errors like Type I (false positives) and Type II (false negatives) can undermine conclusions, emphasising the need for robust designs and appropriate statistical power, typically set at 0.80 to balance error minimisation and resource efficiency. Low power, often due to small sample sizes or weak effects, risks unreliable results but can be mitigated through larger samples, stronger manipulations, or within-subjects designs. Additionally, challenges like publication bias and *p*-hacking distort research integrity, necessitating solutions such as replication, registered reports, and transparent data sharing to ensure unbiased findings.

Despite its centrality, NHST faces criticisms for misinterpretations, particularly around *p*-values, and for its reliance on rigid significance thresholds like $p < 0.05$, which often dismiss nearly significant results and underreport non-significant findings. Critics also argue that NHST provides limited insights into effect strength or practical implications, suggesting alternatives like Bayesian statistics or complementary practices such as reporting effect sizes and confidence intervals. These enhancements offer richer insights and address the limitations of NHST, which

remains widely used but is increasingly supplemented by innovative approaches to ensure rigorous and meaningful research outcome

VERSION HISTORY

This page provides a record of edits and changes made to this book since its initial publication in the JCU Open eBooks Collection. Whenever edits or updates are made in the text, we provide a record and description of those changes here. If the change is minor, the version number increases by 0.1. If the edits involve substantial updates, the version number increases to the next full number.

The downloadable files available for this eBook always reflect the most recent version. We are always seeking to improve our open eBooks. Please submit your suggestions, recommendations, and/or corrections here.

Version	**Date**	**Details**
1.0	12 August 2025	This ebook published on JCU Open eBooks platform
2.0		

ATTRIBUTION AND ACCESSIBILITY INFORMATION

The JCU Library Open eBooks program advocates that education should be available to everyone, which means supporting the creation of free, open, and accessible educational resources. We are actively committed to increasing the accessibility and usability of the open texts and resources we produce.

Accessibility Features of the Web Version of this Resource

The web version of this resource has been designed with accessibility in mind and incorporates the following features:

- Designed to consider the needs of people who use screen reading technology.
  - All content can be navigated using a keyboard.
  - Links, headings, and tables are formatted to work with screen readers.
  - Images have alt text.
- Information is not conveyed by colour alone.

Accessibility Assessment

Below is a short accessibility assessment of key areas that have been assessed during the production process of this open text. The checklist has been drawn from the Open Publishing Guide for Authors by the University of Southern Queensland. While a checklist such as this is just one part of a holistic approach to accessibility, it is one way to begin our work on embedding good accessibility practices in the eBooks we support.

We hope that by being transparent on our current books, we can begin the process of making sure accessibility is top of mind for all authors, adopters, students and contributors of all kinds on all our open-text projects. As such, we welcome any feedback from students, instructors or others who encounter the book and identify an issue that needs resolving.

Accessibility Checklist

Category	Item	Status
Organising Content	Content is organised under headings and subheadings	Yes
Organising Content	Headings and subheadings are used sequentially (e.g. Heading 1, Heading 2, etc.)	Yes
Images	Images that convey information include Alternative Text (alt-text) descriptions of the image's content or function	Yes
Tables	Tables include column headers and row headers where appropriate	Yes
Tables	Tables include a title or caption	Yes
Tables	Tables have adequate cell padding	Yes
Weblinks	External web links open in a new tab. Internal web links open the same window.	Yes

Accessibility Improvements

While we strive to ensure this resource is as accessible and usable as possible, we might not always get it right. We are always looking for ways to make our resources more accessible. If you have problems accessing this resource, please contact the JCU Library Open Education team to let us know so we can fix the issue.

You can contact us via Web form: Submit a question to JCU Library

Third-Party Content

In some cases, our Open Educational Resources include third-party content. In these cases, it may not be possible to ensure the accessibility of this content.

Attribution

This book *Critical Thinking in Psychology: Dispositions, Cognitive Insights, and Research*

A number of chapters of this book are gratefully attributed as being adaptations from each of the following sources:

- *How to Think Critically* by Stephanie Gibbons & Justine Kingsbury, University of Waikato, is used under a CC BY-NC Licence.
- *Neuroscience, Psychology and Conflict Management* by Judith Rafferty, James Cook University, is used under a CC BY-NC Licence.
- *Mastering Thinking: Reasoning, Psychology, and Scientific Methods* by Michael Ireland, University of Southern Queensland, is used under a CC BY-SA Licence.
- *Research Methods in Psychology* by Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler & Dana C. Leighton, Kwantlen Polytechnic University, is used under a CC BY-NC-SA Licence.

Please note: Individual chapters, images, videos, animations and activities may have different licences applied. Check the licence for all chapters and materials and contact the copyright holders for express permission if you wish to use them outside of the conditions stated.

This statement was last updated on 12 August 2025

Accessibility Information Attribution

The accessibility information overview has been adapted in parts from:

Open Publishing Guide for Authors by the University of Southern Queensland. Licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence.

ABOUT JAMES COOK UNIVERSITY

For over 50 years, JCU has empowered graduates with the education and experience needed to thrive in a global workforce. We take pride in serving regional and remote communities, and remain committed to offering study experiences that are exciting, challenging, and rewarding. As Australia's only university established to address the unique issues of Northern Australia and the Tropics worldwide, JCU is dedicated to fostering prosperity in tropical regions across the globe.

Study at JCU and unlock your potential with world-class education, cutting-edge research, and a vibrant campus life that inspires connections and lifelong memories.