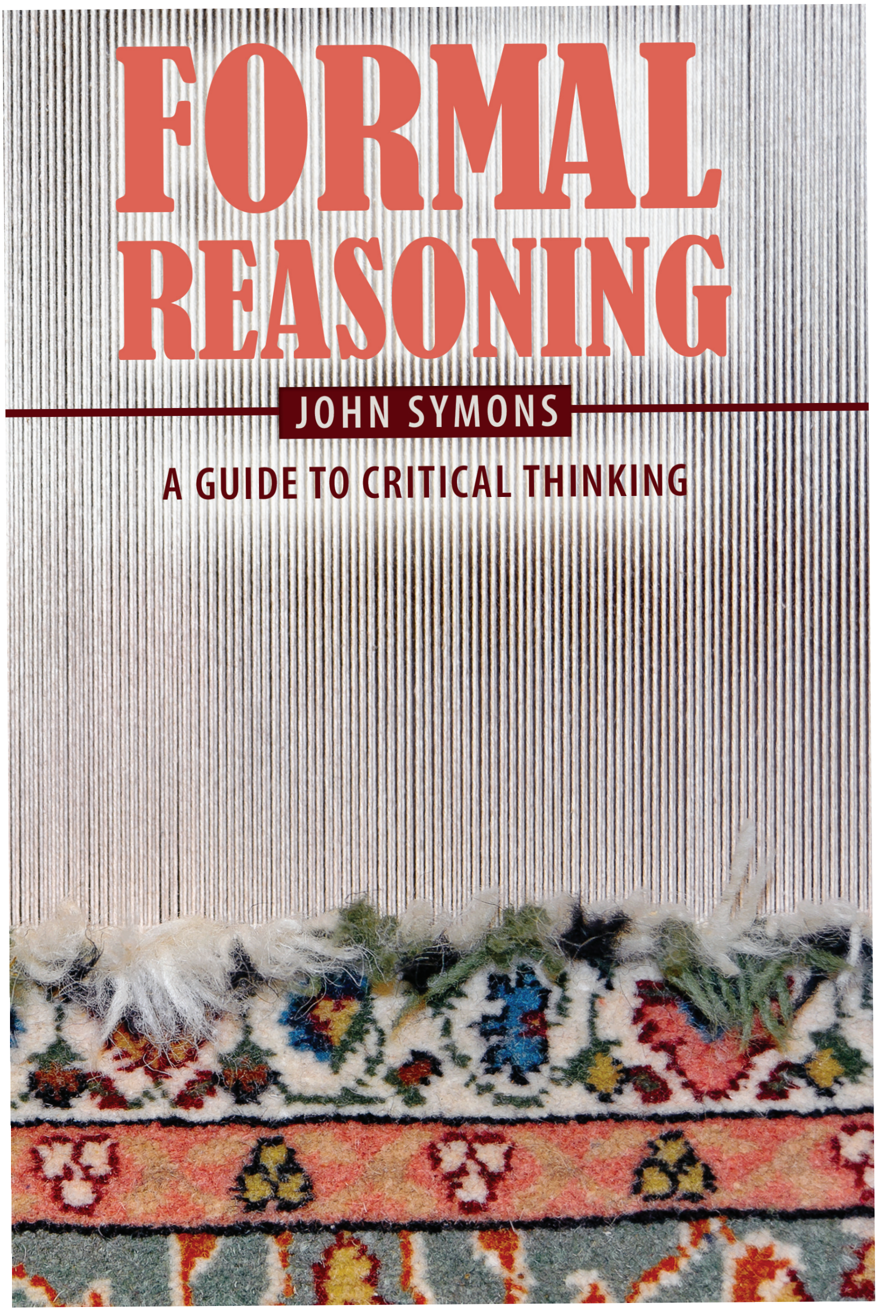


Kendall Hunt
publishing company

Cover image © Shutterstock.com

Kendall Hunt
publishing company

www.kendallhunt.com
Send all inquiries to:
4050 Westmark Drive
Dubuque, IA 52004-1840

ISBN 978-1-5249-2221-4

Published in the United States of America

CONTENTS

ACKNOWLEDGMENTS

It has brought me great joy to teach logic and critical thinking over the years to students at The University of Texas at El Paso and The University of Kansas. I have found it deeply rewarding to help open the door to the beautiful worlds of mathematics, philosophy, and computer science to students who previously considered themselves *math-phobic* or had some other self-undermining prejudice. I am grateful to my students and to my employers; the taxpayers of Kansas and Texas for the opportunity to teach fundamental critical thinking skills in an age where adherence to the norms of good reasoning appears to be in decline.

I have been blessed with great teachers, colleagues and graduate students. I hope that this book reflects some of their influence. Lester Reiss, Jaakko Hintikka, and Akihiro Kanamori taught me logic and how to teach logic. None of them would agree with everything in this book and indeed each of them would find something to which they would strongly object. Hintikka's view of logic has had the greatest influence on my teaching. Specifically, his emphasis on the ameliorative role of epistemology and logic shapes the entire project of this book. While he would have criticized the reliance on psychology and evolutionary explanations of cognitive bias, he was committed to teaching logic as a way of improving the strategic reasoning skills of students and I have tried to do the same.

This book is dedicated to my wife Irina Symons. Living with Irina, a philosopher who is as intellectually uncompromising and candid as anyone I know, has forced me to think more critically and creatively in all respects. Her criticism shaped large parts of this book, her love and intelligence shape our life together.

PREFACE

After reading this book, I hope that you will share my view that mastering the basic elements of formal reasoning is a practical way of overcoming bias, thinking clearly, and improving the quality of one's decision making. Developing a self-critical desire to achieve excellence in reasoning involves more than simply being exposed to lists of fallacies or some basic sentential and first-order logic. Readers need to understand why they ought to take formal methods seriously, why formal methods can help them to overcome their own limitations, and why failing to reason well can be costly and sometimes even blameworthy.

I hope to provide readers with a book that connects basic logic and probability theory in an accessible and applicable way with our growing understanding of the psychology of reasoning. A related goal I had in writing this book is to open a door into the world of formal reasoning for readers who might otherwise feel alienated from mathematics and from the more formal reaches of philosophy. In addition to learning how to evaluate arguments, readers will be introduced to the practice of giving proofs, and will come to understand what it means to think though problems like a mathematician. I am confident that readers will enjoy the material in this book and that they will be encouraged to explore logic, mathematics, computer science, and other areas of formal reasoning more deeply.

Human beings manage extremely subtle and complex cognitive tasks very well. We are very good at many things. However, our ability to reason is systematically limited in ways that have been illuminated by experimental evidence from disciplines such as psychology and economics. We humans are finite beings with deeply ingrained habits of thought. We are now in the fortunate position of having empirical evidence that illuminates many important aspects of our own reasoning.

Thanks to the work of twentieth century logicians and philosophers, we are also fortunate to have a highly developed theoretical understanding of what good reasoning looks like. Our understanding of what a proof is, or how to evaluate the validity of an argument is better than it has ever been. While the empirical sciences provide our best access to how it is that we *actually* reason, philosophy and logic have helped us to understand how we *ought* to reason.

Most hardworking readers can quickly develop the ability to judge when reasoning is going wrong. Once we can tell the differences between excellent and defective reasoning, the next step is to learn strategies that help us to become excellent thinkers for ourselves. The strategies presented in this book are drawn from formal philosophy, logic, mathematics, cognitive psychology, and economics. Together, these areas of research can help us to understand the patterns of reasoning that feature commonly in ordinary life and how we can improve our ability to reason. This book organizes the results of this research in an accessible and useful way.

The most prominent and widely used introductory logic textbooks have all had a similar look and feel with a great deal of overlapping content. There is only so much that can be written about sentential and first-order logic, so perhaps the homogeneity of these texts is to be expected. At the same time, much of the choice of content in these textbooks can be explained by inertia. Publishers have encouraged authors to provide textbooks that fit the standard ways that philosophers have been teaching logic over the decades. However, at a certain point, teachers of logic and critical thinking must reassess the content of these traditional courses in light of our changing understanding of the subject matter. I believe that the time to rethink the content of our logic courses is long overdue.

Much of the content of traditional introductory logic books is dropped from this book in order to make room for what I regard as more useful or important material. I completely omit syllogistic logic. I make no excuse for skipping it. Nor do I feel any strong need to defend my exclusion of other standard pieces of the repertoire of logic textbooks such as Mill's methods and analogical reasoning. In practice, omitting some traditional topics allows room in the semester for some discussion of the psychology of reasoning and for an introduction to probability theory. No introduction to critical thinking in our time should go without a healthy dose of Tversky and Kahneman's work in the mix.

After teaching logic and critical thinking for nearly two decades at a variety of institutions, my approach to the subject has changed. Like many philosophers, my approach to teaching logic was shaped by the influence of W. V. Quine's *Methods of Logic* and Herbert Enderton's *A Mathematical Introduction to Logic.* In addition to Quine and Enderton, I was strongly influenced by the goals that Jaakko Hintikka and James Bachman describe in their book *What If?: Toward Excellence in Reasoning*. Hintikka and Bachman were motivated by the desire to cultivate the strategic reasoning abilities of their readers. They emphasize the role of questioning and imagination in ways that logic textbooks had failed to do prior to their work. They saw education in logic as the cultivation of excellence in reasoning rather than the inculcation of rules and prohibitions. My approach to teaching at the introductory level continues to be motivated by the same principles that Hintikka and Bachman championed. These books shaped my early approach to the subject and there are still some traces of my early courses in the current book.

The great tradition in modern logic stretching from Gottlob Frege to the present continues to be a cornerstone of philosophical research in our time. I believe that no philosophy student should leave an undergraduate program without at least some appreciation and a basic understanding of polyadic quantification theory. Like most logic textbooks, this one provides an introduction to both sentential and first-order logic in a relatively conventional, albeit stripped-down manner. I hope that teachers will feel that my introduction of the quantifiers is accessible enough for inclusion in critical thinking courses that might otherwise avoid first-order logic entirely.

The pedagogical goals associated with the teaching of critical thinking are notoriously difficult to achieve and many of the standard strategies fail.[1] Courses in logic and critical thinking have long been

[1] Willingham, Daniel T. "Critical thinking: Why is it so hard to teach?." *Arts Education Policy Review* 109.4 (2008): 21–32.

the purview of philosophers and thus philosophers deserve criticism for failing to develop useful strategies for motivating and educating our fellow citizens to become effective critical thinkers. Philosophers generally prefer thinking about philosophical problems to helping other people think more clearly. We have been satisfied with our work as educators in part because we notice how well philosophy majors perform relative to other students. We can proudly point to evidence that philosophy majors perform better on standardized tests of reasoning ability than most of their peers.[2] Perhaps, in part this is the result of thinking carefully about fundamental and abstract problems and we philosophy teachers can take some credit—more likely, the excellent performance of philosophy students on tests that measure critical thinking is due to selection bias.

Philosophers have generally not kept track of the pedagogy of critical thinking and as a result other disciplines have begun to encroach on our traditional territory. In recent years, some have argued for an approach known as critical thinking across the curriculum.[3] Proponents assert that the standard philosophical approach to the teaching of critical thinking is too abstract and that the relevant skills can and should be taught in conjunction with the content of courses in all disciplines. They argue that critical thinking skills cannot be taught in isolation but must be grounded in students' understanding of content. While it is true that under certain circumstances, it is easier to think more critically about subject matter we understand well, the trend toward critical thinking across the curriculum is problematic for a number of reasons.

To begin with, it ignores the time and effort involved in a proper education in critical thinking. Students need to practice in order to develop these skills and this practice takes valuable time. Even more importantly, students need instructors who can demonstrate excellence in those skills and who can guide them in their practice. Teaching critical thinking is a labor-intensive and interactive enterprise. It cannot be

[2] Comparisons of performance on standardized tests organized by college major can be found here http://dailynous.com/value-of-philosophy/charts-and-graphs/ (last accessed November 11, 2016)

[3] Chaffee, John. "Teaching critical thinking across the curriculum." *New Directions for Community Colleges* 1992.77 (1992): 25–35.

replaced by the passive consumption of PowerPoint presentations and the regurgitation of lists. In order for a course in some content area, say, for example, history or engineering to provide a meaningful education in critical thinking, instructors must sacrifice significant space in the syllabus and time in the classroom.

Presumably, we should not expect every course in the university to teach basic critical thinking—to explain basic notions such as premise, inference, conclusion, and fallacy. Nor do we expect a wide range of courses to teach students how one can actually evaluate the validity of arguments. Regrettably, many of the pronouncements concerning critical thinking across the curriculum that we find in print or that we hear in university committee meetings, indicate that proponents have a relatively vague grasp of the skills and content associated with critical thinking. For example, in the contemporary university, critical thinking is often confused with ideologically motivated critiques of various kinds. Having a high-IQ, being a political radical, or knowing a great deal about electrical engineering or Victorian literature does not necessarily make one a critical thinker. We must expect instructors of critical thinking to excel as critical thinkers themselves. Since college faculty members generally assume that they themselves are good thinkers, they assume that they can easily instruct their students in critical thinking. We should not share this assumption. In practical terms, it is likely that we will continue to recognize the need for a single course in the university curriculum where students will be educated in the basic skills that we associate with critical thinking. I am not arguing that critical thinking is the sole purview of philosophers, far from it. In fact, this book is informed as much by economics, mathematics, and psychology as it is by philosophy. However, for now, philosophers continue to be best placed to teach this kind of course.

Admittedly, philosophers have generally done less than we should have to ensure that our courses succeed in educating our students in critical thinking. We have certainly failed to change our teaching in ways that take account of advances in the scientific understanding of reasoning. Vestiges of the traditional antipsychologism that marked early analytic philosophy linger in the way we teach logic and critical thinking. Standard introductory courses in formal and informal logic are usually conducted without any mention of the psychology of reasoning and decision making. This omission is especially glaring when it comes to the scientific literature on bias and heuristics. In my view, ignoring the fact of bias in reasoning is a wasted opportunity for logicians and philosophers who teach introductory courses. It is precisely the debiasing effect of logic and probability theory that makes formal methods relevant and useful in the development of critical thinking.[4] In order to recognize this salutary effect, students need to first understand their own limitations as reasoners. Taking account of the growing scientific literature on cognitive bias and other limitations, and recognizing that ordinary common sense is the starting point of our enterprise allows the logician to demonstrate the value of logic and formal reasoning in a way that is both convincing and practical.

Most adults will acknowledge that we are not always effective or reliable thinkers and decision makers. We are aware that feelings are not good guides to reasoning, we understand that problems should be approached from a variety of perspectives, and we recognize that there are better and worse ways to evaluate arguments and evidence. Our problem is to know how to live up to these general principles in practice. Critical thinking is not easy and the first step is to begin with a healthy measure of self-criticism. If one is satisfied with one's ability as a critical thinker (which too many of us are) one can simply compare oneself now to oneself at different points in time. Even the most confident of us will acknowledge, for example, that there are some times in our lives when we are better or worse thinkers and decision makers. If we are tired, angry, afraid, hungry, have been drinking alcohol, or are in the grip of some ideology or dogma we are less reliable thinkers than when we are well-rested, calm, sober, and open-minded. Obviously, not all reasoning is equally effective. Acknowledging that there are better and worse ways of reasoning and that we human beings are fallible is a great place to start.

Genuine education involves the effort to become a better self. It requires humility insofar as one must first realize one's own limitations. It also requires the confidence and ambition to recognize one's own capacity to go beyond one's current limits.

[4] For a historically informed account of the debiasing role of formalism, see Novaes, Catarina Dutilh. *Formal languages in logic: A philosophical and cognitive analysis*. Cambridge University Press, 2012.

1 Why Should You Think for Yourself?

Our decisions are more likely to reflect our values and further our interests if we have developed basic critical thinking skills. A critical thinker is more likely than their uncritical neighbor to make better decisions when it comes to health, relationships, parenting, finance, scientific inquiry, and the like. Critical thinkers fare better in life insofar as they more easily recognize bias, bad arguments, scams, prejudice, propaganda, and superstition.

Critical thinkers are better able to avoid being trapped in bad habits of thought, to cultivate their imaginations, to more consistently follow their principles, and to live in truthful and honest ways. By contrast, uncritical thinkers can easily sink into self-delusion, conformism, contradiction, confusion, and wishful thinking. This book is intended to help readers become better critical thinkers and decision makers.

One of the most serious obstacles to thinking critically is the problem of confusing facts about our feelings with reasons for holding beliefs. Wishful thinking is the most prominent and challenging example of this problem. After carefully examining wishful thinking, this chapter presents moral and practical reasons for developing one's ability to think critically.

1.1 Wishful Thinking

Critical thinking does not provide a path to cozy and reassuring beliefs. It will not necessarily support your favorite ideology, it is potentially disruptive to some aspects of your current way of life, and it may even irritate some of your friends and family. Nevertheless, a critical thinker should favor truth over comfort. We ought to favor truths even though we sometimes derive some pleasure from believing falsehoods. Typically, careful students of critical thinking find they must abandon at least some of their cherished opinions or comfortable habits of thought. Doing so requires courage, intellectual maturity, and humility. Not all of us can be courageous and mature all of the time. However, an education in critical thinking requires that, at a minimum, you aspire to these virtues. Some people claim that they would rather be wrong and feel good than be right and not feel good. This book is not for them.

Let's begin with the problem of wishful thinking. Wishful thinking results from confusing one's feelings about some state of affairs *S* with reasons for believing that *S* is the case. We will define wishful thinking more explicitly below, but first consider an example: Imagine an unattractive person who takes some pleasure in believing that attractive people are defective in some way. Perhaps our wishful thinker might believe that attractive people are more superficial, or that they are generally less intelligent than unattractive people. The fact that someone finds a belief enjoyable or comforting is not a good reason for them to find it plausible. The wishful thinker may take some consolation in the false belief that attractive people are stupid or superficial. This is an instance of wishful thinking and it can lead to negative consequences.

Wishful thinking can interfere with our decision making in a variety of ways. Imagine how this person's interactions and attitudes will be influenced by their false belief. One can predict that this wishful thinker might behave in ways that are unfair, might miss opportunities for friendships, and might otherwise harm, or be harmed as a result of believing a comforting prejudice.

While it is probably false that attractive people are significantly more superficial or stupider than the average person, it *might* be true. How would we figure out whether it is true? The best way would be to conduct the most reliable empirical studies that we could. Presumably we could take a representative sample of attractive people from the general population and subject them to tests that would allow us to determine their intelligence. Wishful thinkers are rarely interested in running experiments of this kind. Generally speaking, wishful thinkers are not likely to have consulted the best evidence available. Instead, our wishful thinker, for example, simply assumes that their belief concerning attractive people is true. Their reasons for belief have little to do with an unbiased consideration of evidence. If challenged, they are likely to search their memory for episodes where they met stupid or superficial attractive persons. The wishful thinker regards such memories or anecdotes that they have heard from others as evidence that supports their prejudice. Of course, memories or anecdotes that do not support their prejudice tend to be ignored.

One of the most serious problems for the wishful thinker is their tendency to assess all evidence through the filter of their prejudices. In this way, wishful thinkers are especially vulnerable to what is known as **confirmation bias**. In later chapters, we will revisit confirmation bias in detail. For now, it suffices to say that confirmation bias is the tendency to favor evidence that supports one's cherished beliefs while illegitimately ignoring evidence that does not support those beliefs.

Let's consider another kind of example: Imagine that some desirable career path is unsuited for us in some way. How can we tell when it is time to change course in life and explore new opportunities? Important decisions like this require careful deliberation and wishful thinking in these contexts can prove disastrous. Here is an extreme case:

> *Much as Frank might like to become a professional basketball player, he simply cannot master the skills required for the job. As it becomes clearer to his friends and family that he lacks the required ability, he still holds on to the comforting dream that—in the face of evidence to the contrary—becoming a professional basketball player really is his destiny.*
>
> *At a certain age, the aspiring professional basketball player is a pitiful sight; those around him correctly regard the forty-year-old Frank who spends his time training for a career in the NBA as suffering from very serious delusions.*

Persistence, ambition, and commitment are admirable, but at what point is the aspiring basketball player harmed by his devotion to an unrealistic dream? At what point can we say that he has squandered precious time and energy that might have been more productively spent in other ways?

> ### *Wishful thinking:*
>
> *One can be accused of wishful thinking insofar as one holds a belief because one wishes the belief to be true or because one finds the belief to be pleasing.*

There may be some value or pleasure in holding onto the dream of becoming a professional basketball player. But even in the pursuit of improbable dreams or highly ambitious goals, it is surely better to have a realistic assessment of our prospects than not. It is almost always better to pursue our dreams with an accurate understanding of how likely it is that we will succeed. Our aspiring basketball player Frank is a victim of wishful thinking.

One can be accused of wishful thinking insofar as one holds a belief because one wishes the belief to be true or because one finds the belief to be pleasing. It is common to see friends and family members who, because of wishful thinking, continue romantic relationships or career paths that are harmful to them. It is pleasant to imagine that in spite of their bad behavior, one's romantic partner is a good person *on the inside* and that their bad behavior is changing, or perhaps could be changing. Wishful thinkers in failing relationships might exhibit confirmation bias by overemphasizing the happy or pleasant memories of their partners and underemphasizing memories of bad behavior.

Notice that it might be true that our wishful thinker's romantic partner in fact *is* a good person *on the inside* (whatever that might mean). However, the truth or falsity of this judgment is independent of whether that claim pleases or displeases the wishful thinker. Wishful thinking can keep us believing (sometimes against all the evidence) that our partner lives up to the idealized image that we find so pleasant. We can find a variety of examples of how wishful thinking can keep decision makers from recognizing painful or unpleasant realities, not only in our personal lives, but also in business, and in government.

Common sense tells us that wishful thinking can lead to poor decision making, but it takes a little analysis to see exactly where we go wrong when we engage in wishful thinking? At this point we need to introduce some formalism:

Wishful thinking has the following form:

I wish that *S*
Therefore
S is the case

In this pattern, *S* stands for any assertion about the way the world is. *S* might be true, it might be false, however, the fact that one *wishes* that *S* were true is not a good reason to conclude that *S* is true.

Notice that the problem here is not that the wishful thinker takes pleasure in an idea, or would prefer the world to be one way rather than another. It is perfectly reasonable to wish that your doctor were better than they are, or to prefer that your romantic partner be generous and considerate. *The problem with wishful thinking is that we mistakenly or unthinkingly take our wishes, or preferences to be* ***reasons to believe*** that *S* is the case. The fact that one likes, wants, or prefers some state of affairs *S* does not mean that one is entitled to believe that *S* is the case. Notice that the wishful thinker is at fault because they believe *S* **for a bad reason**.

Wishful thinking can be a particularly disruptive and costly error in reasoning. As we recognize from our own experience and the experiences of others, it can lead to expensive financial mistakes, commitment to failed political ideologies, terrible romantic relationships, loyalty to bad doctors, and countless other harmful patterns in our behavior and beliefs.

1.2 Intellectually Mature Thinkers Try to Distinguish Their Feelings from Their Beliefs

In order to counteract the influence of wishful thinking, we need to carefully distinguish our feelings and our preferences from our beliefs. A characteristic of wishful thinkers is that they are unable to distinguish their feelings about some belief from the reasons to accept or reject the truth of a belief.

The following statements are assertions concerning matters of fact. There are good reasons to believe that these assertions are false. Nevertheless, they are the kinds of platitudes that we frequently hear in conversations:

> "What goes around comes around."
>
> "Celebrities are all miserable."
>
> "Everybody has something good in them."

We have no evidence to believe that all celebrities are miserable. In fact, the existence of one nonmiserable celebrity is enough to make the claim false. But why might we be inclined to believe that something like this is true? Quite simply because if it were true, it would offer those of us who are not celebrities a pleasant kind of consolation in our lower social status. Notice that these claims are not, in themselves, examples of wishful thinking. In themselves, these are simply three claims about reality.

Notice that if we imagine what **genuinely good reasons to assent** might look like in these or any other cases; we would not find our *desire for the claim to be true* among them. Perhaps it is true, for example, that everyone has something good in them. It has some plausibility insofar as it is vague enough to be possibly true., For example, a serial killer or a dictator might be a terrible person, but maybe their mothers recognized something good about them or maybe they are kind to animals—they might really like kittens. Given the vague nature of the claim that everybody has something good in them, it is relatively easy to provide good reasons for believing that this is true. The vagueness of the claim is also evidence that it is not really telling us very much; even if true, it is not a very informative claim.

One could imagine having reasons, independently of what one wishes or prefers to believe any of these claims. Take the first "What goes around, comes around." Perhaps one lives in a traditional society with limited access to scientific institutions and methods where one relies on experts in religious or philosophical matters to construct theories about moral or metaphysical matters. "What goes around comes around" is a crude version of the classical Indian philosophical doctrine of Karma. If one relied on philosophers and religious figures in such matters, one might take their pronouncements as reason to believe that Karma is a fundamental principle governing the natural world. You believe this, for instance, because experts on such matters tell you that this is the case and perhaps you have never seriously questioned how they could know such things. Just as I have never bothered to question experts on the difference between beech trees and elm trees, a person who lived in the Mauryan Dynasty (from 322 to 185 BCE) in India might see no reason not to trust what he or she imagines are experts in religious matters. If this is the case, then the belief in Karma is not an instance of wishful thinking. Your belief in Karma is not necessarily supported by good reasons, but *it is supported by something other than your feelings about Karma*.

We often hold on to ideas without caring about whether they are true or whether we have good reasons for believing them. Most of the time, this is relatively harmless. In most of the examples above, there is relatively little at stake. For example, the belief that celebrities are miserable or shallow, will play no role in any significant decision that most of us will make. Our beliefs about the private lives of stars, like so many of our beliefs are simply not relevant to decisions that really matter. In fact, while it is certainly better to avoid falsehood and arbitrariness in our beliefs, it would be a waste of time and energy to care too much about all of our beliefs. We do not have unlimited time or energy and the vast majority of our beliefs and opinions are simply not important enough to worry about.

Given that most of us spend relatively little time determining the truth regarding the vast majority of our beliefs, it is important not to be too confident. Happily, we do not in fact need to be genuinely confident most of the time because the cost of having false beliefs about most matters is relatively low.

A critical thinker need not aim for epistemic sainthood. However, some of our beliefs play a crucial role in decisions that have consequences for our own well-being or for the well-being of others. In these cases, not caring about the truth, and being unable to recognize the role of wishful thinking in our reasoning would be blameworthy.

There are, presumably, some people who do not care about the harmful effects of their decisions on themselves or others. Such people have the luxury of believing what they prefer to believe simply because they prefer to believe it. We are familiar with people for whom truth and the well-being of others is not relevant to their decision making. Such people are dangerous to others and should not be trusted with the power to make important decisions.

What wishful thinking is versus why it can be problematic:

Wishful thinking is holding beliefs solely or primarily because you wish they were true, independently of the truth or falsity of those beliefs.

Notice that wishful thinkers **can believe true** things for a bad reason, namely because they wish they were true.

Wishful thinking **poses practical problems** because it disposes one to believe false things because one wishes they were true. Believing false things can lead to practical and moral problems.

Part of what it means to become intellectual mature is to care about the difference between true and false claims and to become open to correction. Reasonable and intellectually mature thinkers are willing to change their minds, for example, by being willing to abandon an opinion or a belief when there are good reasons to do so. Intellectually immature thinkers do not concern themselves with evidence or having good reasons. Instead, they can make themselves invulnerable to rational persuasion by simply ignoring it. By contrast, a reasonable person is open to rational persuasion. In this book, we will explore the idea of rational persuasion and will carefully distinguish it from irrational or manipulative forms of persuasion.

Wishful thinking is an example of the kind of self-undermining habits of thought that we will examine in more detail in this book. It is a particularly pernicious bias that is very difficult to overcome. I start with it here because it is easy to see how it can have moral as well as cognitive consequences.

Intellectually mature thinkers avoid wishful thinking by:

- Distinguishing reasons for beliefs from their feeling about beliefs
- Caring about the difference between true and false claims
- Being willing to change their minds and by being open to correction
- Being able to tell the difference between rational persuasion and manipulation

1.3 Why It Is Sometimes Morally Wrong to Think Uncritically?

Many intelligent people embrace homeopathic medicine, astrology, palm reading, belief in ghosts, and the like. For the most part, they believe such things for reasons that have nothing to do with the careful evaluation of evidence. Perhaps they believe in such things because people they are affiliated with (e.g., their friends, families, coworkers, or members of their political party) say that they believe them. Perhaps they say they believe these things because doing so is comforting or entertaining. Perhaps they simply have never had a reason or an opportunity to think carefully about the claims of homeopaths or astrologers. Alternatively, perhaps they have decided that the costs of thinking critically about such things outweigh the benefits of having true beliefs about these matters.

Why might an otherwise well-informed and reasonable person claim to believe that astrology provides useful knowledge of the future? There are a variety of reasons that some people believe in astrology, some of which we will discuss later in this book. Most of us simply have not given the question much thought and simply treat astrology as an amusing recreational activity. Most of us are unlikely to adopt an aggressively skeptical attitude to astrology in a normal social setting because doing so might come at some social cost. Perhaps denying the claims of astrologers might alienate one's romantic partner, one's family, or one's broader community. In cases like this, one might determine that the social cost of being critical of astronomy is too high and that it is worth staying quiet. Perhaps there has never been a decision in one's life that has involved astrology that is sufficiently important to merit serious reflection on the merits of astrology. In this case, the question of whether astronomy is a reliable source of evidence has simply never arisen. In any event, it is usually relatively cheap, harmless, socially acceptable, and perhaps even fun to be irrational about a wide range of matters, including astrology.

It is worth distinguishing the kind of assent that we give to beliefs for reasons of social convenience from genuine assent. We would likely pay very high prices for the services of astrologers if we really believed that they had access to facts about the future. The fact that fortune-tellers and astrologers charge relatively modest sums is an indication that the market value of their service is a function of their role as entertainment rather than the provision of special access to the future. If price is any indication, then even their clients do not really believe in fortune-tellers.

Most of us do not really believe in astrology, but instead treat it as a kind of amusing recreational activity. We can recognize as a matter of common sense that astrology has no real predictive power. Why?

- Contrary to the predictions of astrology, twins usually experience different events during their lives.
- For the most part, astrologers are not very well paid. If they were able to accomplish what they promise, they would be very well paid indeed.
- Large numbers of people with different birthdates, for example, those killed during the attack on the World Trade Center on

9/11 or the victims of a plane crash often suffer the same fate at the same time while having very different horoscopes.

- How could the gravitational effects of stars and planets be more relevant to our fates than the gravitational effects of nearby objects?
- Why are only a small subset of stars and the local planets considered relevant to our fate?
- Astrological predictions are often wildly contradictory, consider a set of horoscopes from the same day from different sources. They usually disagree.

Common sense tells us that astrology is not to be trusted. However, as we saw above we tolerate irrationality in some contexts because of other values. Irrationality is often fun and socially useful, so why then would one wish to embark on the costly and potentially upsetting path of reasoning clearly? To convince you, I must draw on the resources of moral philosophy rather than logic for assistance. This is because logic and other formal methods by themselves will not give you reasons to be rational; for example, the techniques you will learn in this book do not include reasons why you ought to employ these techniques. However, there are clear cases where being a responsible thinker and decision maker is morally required of us. If the irrational person cares about their moral obligations (and such a person might not care), one can point to situations where deliberately failing to reason well is unethical. For example, when making decisions that significantly impinge on the well-being of others, it is clearly wrong to opt for irrationality, ignorance, or prejudice.

> Think of cases where decisions involve the possibility of serious harm to other people. In these cases, if it is within one's power to reason responsibly about these decisions, then one *ought* do so.

Imagine, for example, not thinking responsibly about how best to care for an aging parent, whether to vaccinate your children, contribute money to charity, drive at high speed, and so on. Would it be morally acceptable simply to toss a coin in order to make important decisions like these? No. In cases like these, we are ethically obliged to reason carefully. We can be (and often are) held morally and sometimes even legally responsible for negligent decision making. Whether we like it or not, in cases where our decisions have significant consequences for others, deliberate failure to think carefully is blameworthy.

There are instances where treating a child with homeopathic medicine rather than real medicine would be a genuine harm to the child. Parents who fail to think critically about the medical care of their children are failing in their moral obligation to avoid harming their child unnecessarily. This is because **homeopathic medicine is not effective in treating any illness**. At this point, readers are invited to do a little research on homeopathy for themselves.

Perhaps one does not care about one's moral responsibilities to others—one might be the kind of unpleasant character who is concerned only with their own well-being. As we will see in the next section, even immoral thinkers have a reason to pursue critical thinking insofar as it allows selfish souls to pursue their own interests without being manipulated by others.

1.4 Resisting Social Conformism, Propaganda, and Commercial Advertising

A basic ability to think critically is indispensable for all of us who hope for some degree of autonomy with respect to the influence of social media, commerce, government, and popular culture on our preferences and on our decision making. The pressures of social conformism and marketing are pervasive and pernicious aspects of contemporary life. It is a simple fact that advertisers and producers of mass culture are motivated primarily by profit and do not have our best interests at heart. Since this circumstance is unlikely to change in the near future, critical thinking can help us, on an individual level at least, resist the persuasive techniques of those who see us as consumers to be manipulated and separated from our wealth.

Commercial interests manipulate consumers through shaping their preexisting desires and anxieties and capturing and directing their attention. This is done through sophisticated uses of psychology and in our time through increasingly personalized use of data about the characteristics and past behavior of individual consumers. We voluntarily transmit a great deal of personal information and behavioral data through our use of social media, and this can serve as the basis of targeted strategies that are optimized to influence the way we spend our money.

Much of contemporary commerce and media is devoted to the study and control of our attention. It does so in ways that leave little space for calm critical reflection. Our screens cultivate a sense of urgency; a need to "check," they create a sense that we might be missing out on something that others have. Our screens feed us an intermittent drip of affirmation, and a constant promise of novelty.

While the all-pervasive commercial culture of our time might have excellent consequences for the broader economy, advertisers and public relations experts work to serve their corporate clients. Thus, the challenge for any person who values their autonomy is to find a way to protect and pursue one's own independent preferences. In our time, the art and science of persuading an audience to accept some claim or to pursue some course of action is usually assigned to a discipline called ***rhetoric***. Modern logicians and philosophers are correctly wary of rhetorical techniques insofar as they can easily be used for unethical political and commercial purposes. In general, commercial advertising and political propaganda is crafted to exploit our weaknesses, causing us to make decisions based on defective reasoning.

For many of us, the greatest luxury is to find the strength to control our own attention and to find a refuge where we have the time to think carefully without being pushed and pulled by commercial interests in the general culture and the conformist tendencies of social media. Education in critical thinking offers us some capacity to achieve an intellectually mature relationship to our own emotional and cognitive life. It also offers us the opportunity to reinforce our autonomy as thinkers and decision makers in the face of a challenging cultural environment.

Why do I want what I seem to want?

It is worth pausing to consider how many of our dreams and desires were cultivated in us by commercial interests. It is sometimes difficult to know what to want, other than what is being marketed to us. Such influences are problematic insofar as those who shape commercial advertising and political propaganda are extremely sophisticated and generally act in their own interests rather than for your benefit. The influence of commercial culture is so powerful that it is often difficult for us to even know how to make independent judgments about our desires. For most of us, it can be difficult to imagine what our preferences, hopes, and aspirations would look in the absence of unrelenting commercial manipulation of our attention and our desires.

The creators of commercial culture are willing to deploy sophisticated psychological techniques to manipulate our attention and to modify our attitudes and behavior. For commercial interests, the most important virtue of a piece of advertising is its persuasiveness. Advertisements can successfully persuade us to accept a claim or to take a particular course of action for bad reasons.

In commercial and political decision making, rhetorical trickery can manipulate us very effectively.

We can be manipulated because of our biases, laziness, strong associations, habits, or limitations on our cognitive capacities.

Advertisers and unscrupulous politicians rely on the fact that we can be flattered, cajoled, and dazzled into thinking and acting against our own interests.

A successful advertisement is carefully crafted to cause us to experience new desires or to feel anxious or uncertain about ourselves in ways that can be exploited for commercial purposes. Advertisements can also manipulate our desire to affiliate with a community of some kind. We are encouraged to signal our loyalty, reliability, or attractiveness to our affiliates in a variety of ways. Our desire to indicate status relative to others is also an important weakness that advertisers can manipulate.

At their most basic level, advertisements work by creating psychological associations between the product or service that they are working to sell and some unrelated good thing. Advertisers call this technique *transfer*. For example, a product like breakfast cereal or lawn furniture might be associated with a sports hero or some other celebrity. An advertiser might associate a car with a sexually attractive person; a political candidate might be associated with patriotic symbols and colors, or a piece of electronic equipment might be associated with high social status or virtues like creativity or coolness. Advertisers anticipate that our positive feelings for celebrities, patriotic symbols, beautiful pieces of music, or beautiful people will be associated with their product.

A positive association need not be based in anything real about the product in order for it to be effective. If questioned, few adults will say they believe that rare athletic prowess results from eating bland breakfast cereal. However, for the purpose of increasing the likelihood that you will buy the product, it is only necessary that the positive feelings you have about some piece of music, celebrity, patriotic symbol, or beautiful person be associated on some level with the product or service being advertised.

The mere fact that a breakfast cereal is pictured together with an admirable athlete clearly should not lead us to think that the cereal has any of his or her admirable qualities.

We can protect ourselves from misleading associations and rhetorical manipulation using critical thinking skills.

The power and ubiquity of advertising and political propaganda in contemporary culture is a very good reason to study critical thinking and logic. Doing so will help you to become more sensitive to manipulation and more independent as a decision maker.

1.5 Common Sense as a Starting Point

In order to improve as thinkers and decision makers, we must first recognize that our ordinary ways of thinking and deciding are subject to manipulation by others, that they are limited, and in some contexts, highly unreliable. As we shall see in later chapters, many of our ordinary habits of thought lead us to make mistakes in ways that are now well understood and predictable. Once we know that we are subject to error in systematic ways, we should hesitate before relying completely on common sense to guide our decision making, at least in important cases. Unfortunately, as we shall see, for many of the challenges facing contemporary decision makers, common sense, by itself, is unreliable and insufficient. Happily, formal reasoning and critical thinking can supplement and correct our commonsense abilities.

Since you are reading this book, it is likely that you are already endowed with a fair share of common sense. But what exactly do we mean by the term *common sense*? Very young children are thought to lack common sense and are encouraged to develop it as they mature. In its contemporary usage "common sense" is a nonscientific term that points to the kind of low-level intelligence and background knowledge that one needs in order to navigate the practical challenges of an ordinary day. We tend to associate common sense with the ability to make simple decisions, to think through practical problems, to detect obviously misleading or confused ways of talking and thinking, and so on. The term *common sense* is not very well-defined. Nevertheless, the concept as we currently use it, points to the basic intelligence that a typical adult human can be assumed to possess. While common sense has significant limits, we have to begin somewhere. While the concept is vague, and the capacity it points to is often unreliable, common sense is the indispensable starting point for developing our skill as thinkers and for cultivating our ability to make well-informed and reasonable decisions.

From *Sensus Communis* to Common Sense

The meaning of the term "common sense" has changed dramatically over the years. In the medieval period, it had a very specific technical role for thinkers who investigated perception and the brain. Medieval thinkers had followed Aristotle in thinking that our experiences of sight, smell, touch, and sound were integrated in one of the hollow spaces between the two hemispheres of the brain. There, a psychological faculty that they called "sensus communis" was thought to allow us to recognize that the delicious taste, the smooth texture, and the purple colored spherical shape that I see are all united in one object; the juicy plum that I am currently enjoying. In this way, all of our senses are integrated by the *common sense*.

Thanks to basic common sense, most adults can recognize some simple cases of faulty reasoning when we encounter them. For example, we notice when someone explicitly asserts two claims that cannot be true together. We call this a **contradiction**. For example:

> "Everyone has the right to privacy but teenagers do not have a right to privacy"

is obviously an unacceptable contradiction. Obvious contradictions jump out to us as clearly objectionable features of reasoning. As we shall see in later chapters, common sense is right to be sensitive to contradiction.

We are also very sensitive to obvious cases where emotion interferes with reasoning. For example, we notice that there is something going wrong in reasoning when someone believes some claim solely because of its **emotional appeal**. Consider the following:

> "I'm sure that my book will get great reviews. Why? It would break my heart to get bad reviews after all the work I put into it, so there's no way it will get bad reviews."

Here, we recognize that the speaker is making a basic error insofar as their emotional state is not a good basis for determining whether or not the book will get good reviews. It might be the case that the book will get bad reviews *and* that the victim's heart *will* break.

We are generally able to tell when bullying or aggression is being employed as a tool of persuasion. There are subtle cases that we are likely to miss, but if someone explicitly argues through **the use of threats or violence**, common sense tells us to reject their reasoning:

> "If you don't agree that Wichita is the capital of Kansas, I will kick you."

In addition to direct threats, obvious attempts to distract or mislead us, or to change the subject, are also objectionable.

Common sense is also able to catch obvious **non sequitur**s. Non sequitur is a Latin term meaning "does not follow" and it indicates an illegitimate jump in reasoning where the ideas that come later have no meaningful relationship with the ideas that preceded them. Common sense can easily detect the problem with saying

> "Lucas is ugly therefore Lucas is smart."

or

> "Oranges are sweet and round, onions are round so onions are sweet."

When we notice simple errors like these, common sense tells us that reasoning is going astray. At this point, prior to learning any formal techniques, it is worth considering each of these examples using common sense alone in order to determine what it is about them that makes them problematic.

Why is a contradiction in an argument a problem? In our first example above, if someone claims that everyone is entitled to privacy while simultaneously claiming that not everyone is entitled to privacy, common sense alerts us to the problem, but what exactly is problematic about it?

There are a number of issues here. On the one hand, from a commonsense perspective, if you assert a contradiction, it is not clear what you are actually asserting. Do you think that everyone has a right to privacy or don't you? If you say both, then which of the two you actually believe cannot be determined.

The second problematic feature of contradictions that common sense finds objectionable concerns actions and decisions. Imagine contradicting oneself while advising someone on some course of action. For example, if a friend advises me that I ought to buy a Jeep and not buy a Jeep, their advice is simply useless. If one believes equally strongly that one ought to perform some action and that one ought not to perform some action, one is not in a position to act in a nonarbitrary manner.

There are some other reasons why contradiction is problematic, but on a commonsense level, the key problem is that contradictions make it hard to tell what the proponent of a contradiction believes or how one should act based on advice that contains a contradiction.

Relatively intuitive cases like these can serve as the starting point for your study of logic—our basic commonsense ability allows us to recognize that there are at least some cases where there is an obvious difference between reasoning well and reasoning badly.

Basic common sense has served our species well over the course of our early evolutionary history. However, the complexity and novelty of the decisions we face in modern life can easily overwhelm common sense. In some modern contexts, it is difficult to make decisions that align with our interests and preferences and we need more than the kind of basic reasoning ability that works reliably in simpler contexts. Our hunter-gatherer ancestors were not faced with the challenge of interpreting the results of medical tests, nor were they required to evaluate the financial risks and rewards associated with debt, insurance, or retirement planning. Many aspects of our personal lives, from coping with social media to managing our finances involve challenges that go well beyond the capacity of unassisted common sense.

In addition to responsibly making complicated decisions in our personal lives, an engaged citizen in a modern democratic society must make decisions concerning complex ethical and political challenges like climate change, the distribution of health care, electronic privacy, the use of military force, or the regulation of the financial industry. When making decisions regarding topics like these we are required to do more than simply follow our gut reactions or trust that common sense can reliably guide us. A responsible citizen should be able to reason well concerning the decisions of their political communities.

It is comfortable and convenient to ignore the responsibility to think carefully about complex problems. Many of us are attracted to simple solutions and are willing to abdicate our duties to think clearly. Simple and emotionally comforting political rhetoric from clever politicians on the left and right of the political spectrum can lull us into obedience by playing to our intellectual laziness and our unwillingness to think critically.

Perhaps you do not believe that you have a responsibility to be a good member of your community. Nevertheless, you will still need to pursue excellence in reasoning in your decision making if you wish to pursue your private preferences effectively. It should trouble you if your opinions on matters that you consider important are not supported by good reasons and evidence and you should be concerned if your decisions about such matters are formed in an arbitrary or poorly-informed manner. This is because your preferences are an expression of what you care about.

It is difficult to imagine what it would mean for someone to not care whether their preferences are satisfied. It seems that you must care about reasoning well given that the inability to reason well can undermine your ability to act in a manner that reflects your preferences.

Of course, there is a sense in which we are free to be foolish and arbitrary in matters that only concern our personal well-being. However, even in cases where decisions only affect us personally, most of us would prefer to make important decisions for good rather than bad reasons.

Exercises for Chapter 1

1. Consider an example of wishful thinking from your own immediate experience. Perhaps you, your friends, or your family members were wishful thinkers at some point. What distinguishes their wishful thinking in this case from healthy optimism?
2. Reflecting on your own beliefs and decision making, in which area of life do you find yourself most guilty of wishful thinking? Career, romantic relationships, friendships, politics, your own abilities, or elsewhere?
3. In "The Recipe" Kendrick Lamar's lyrics include: "If I'm wrong, I don't wanna be right". Can he coherently wish to be wrong? Does he want the world to be as he thinks it is or does he want to be wrong about the way the world is?
4. Given the costs involved in the production and delivery of an advertisement in a mass-market venue, you should assume that most aspects of such advertisements were deliberately included and that they serve the overall purpose of selling the product. Examine an example of an advertisement that you find particularly appealing or successful. Try to determine how the various aspects of the advertisement are working. How do they contribute to its effectiveness?
5. It is worth examining your possessions and asking yourself why you purchased this particular product over an alternative. Do you understand the reasons for your decision to buy this product? For example, did I have a good reason for buying the name-brand pain relief medicine over the generic equivalent?
6. If you are told that there is only one copy of a book remaining for sale on your favorite shopping website does this change your likelihood of buying the book? What if the hotel booking website tells you that there is only one room left "at this price"? Do you trust the website when it tells you these things? How does your perception of scarcity change your interest in buying a product? Should it?

7. While it has important limits, common sense is generally reliable in daily life. Can you identify some of the characteristics of problems where common sense fails to help us make decision?
8. How might a software engineer teach a machine to have common sense? How would you?
9. Compare the role that common sense plays in our decision making with abilities like riding a bike or catching a ball?
10. When you think of people who lack common sense, one familiar image is the awkward, bookish or mathematically-oriented person. What does it mean that we like to think of certain kinds of intelligent people as lacking in common sense? Where do we think they fail? Do these failures tell us something about the nature of common sense?
11. Consider an example of a decision in which it is not important to think carefully? What are some of the characteristics of these decisions?

2 Introducing Arguments

Ideally, before one commits to some belief or course of action one should be in a position to rationally convince oneself that one is justified in doing so. This is a very high standard and we will not always have the time or resources to be completely sure of ourselves on all occasions. It would be great to know that one holds one's beliefs for good reasons but most of the time and with most of our beliefs we are not in a position to be confident. However, in important situations, one should, at the very least, be able to convince oneself that there is *some* reason for one's beliefs and actions and that one is not just acting arbitrarily or irrationally. In other words, one should be able to provide a **justification** or give an **argument** in support of one's beliefs.

> "Because I said so"
> "I just feel like it"
> "That's just my opinion"…
> do not usually count as good justifications

2.1 Arguments

Studying logic and critical thinking is an effective way of strengthening our ability to reason about arguments. The word **"argument"** has a variety of meanings. When logicians and philosophers use it, they mean ***a finite sequence of sentences that is intended to rationally convince an audience to accept some claim.*** Put in this way, we might think that arguments are primarily a matter for academic journals, opinion pages in the newspaper, legal documents, or scientific debates. While these are all venues for argumentation, it is important to recognize that arguments are not just scholarly or legal matters. Arguments are an integral part of ordinary reasoning and decision making. As we shall see, some arguments are good, and should be accepted while others are bad and should be rejected. As mentioned above, one should have good arguments for why one decides one way rather than another.

Not every string of sentences or thoughts is an argument. Among the many strings of sentences that would not normally be considered arguments are:

Directions to the park, shopping lists, recipes for apple pie, descriptions of the aurora borealis, weather forecasts, an explanation of the rising price of oil, and a general's orders to his or her troops.

Normally, an argument will assume some shared common ground between the person giving the argument and their audience. An argument has to begin somewhere and some basic claims have to be assumed. We call these assumptions or starting points the ***premises*** of an argument. The premises are claims that the ***proponent*** of the argument (the person making the argument) takes for granted. Part of being a good thinker is being appropriately cautious before accepting the truth of the proponent's premises. In important matters, one should not simply accept the premises of an argument without having good reasons for doing so. Instead, one should make sure that the premises come from a reliable source or method and are supported by good evidence.

Premises serve as the basis for ***inferences***. Inferences are the sequence of moves that take the audience from the premises to the main contention or **conclusion** of the argument. Inferences move our reasoning along from shared common starting points to steps in an argument and eventually to the acceptance of some contention or conclusion. Some ways of moving from one step to another in our reasoning are good, some are bad.

The proponent of an argument hopes that their audience will be convinced to accept their main contention on the basis of having accepted the truth of the premises and the legitimacy of the inferences contained in the argument. If one accepts the truth of the premises and accepts that the steps are legitimate, then one has no rational basis for rejecting the conclusion of the argument.

Effective critical thinking depends on being able to identify and evaluate the inferences and premises that compose an argument. Without these abilities, one fails as critical thinker, and risks being an irrational decision maker. To see why, consider how closely related arguments are to the practice of giving reasons or of justifying. When we ask for a justification, we are asking a why-question. Let's take a very ordinary case:

Why do you think that we should install solar panels on the roof?

When we ask this question, we are looking for reasons. After all, installing solar panels is an expensive course of action. You can probably think of several reasons why installing solar panels might be a good or a bad idea. In order for someone to be entitled to claim that we should install solar panels, they must justify their recommendation with reasons and reliable evidence. Justifying the claim that one ought to install solar panels on the roof should take the form of an argument that leads from a reasonable set of shared assumptions to the conclusion that putting solar panels on the roof is the right course of action.

Most of us do not want to be arbitrary or careless, especially when it comes to morally important or expensive matters. Thus, it will be important that we check to see whether the beliefs underlying our decisions are justified appropriately. If we are able to give a good argument for why we decided as we did, then we can be confident that we are being responsible decision makers.

The Quality of the Evidence

If the only source of evidence that I am considering is the brochure from the solar panel salesperson, then I am not being responsible about which premises to adopt. The salesperson has an interest in influencing my decision and is likely to present information that influences me in that direction.

The Quality of the Reasoning Based on the Evidence

Even if I have reliable sources of information about the costs and benefits of solar panels, I might still go wrong in my reasoning. For example, my inferences from that information to the conclusion may be flawed in some obvious ways. For example, I might be contradicting myself or my reasoning might exhibit obvious non sequiturs. If so, then I am not being responsible in my reasoning.

2.2 Arguing with Oneself: Reasoning and Decision Making

As we try to decide on the best course of action, our deliberation is frequently unclear and poorly organized. Often, we are driven by a variety of habits, biases, and subconscious factors beyond our immediate control. Frequently, we are distracted by the many demands on our attention and we sometimes lack the time, the energy, the information, or the cognitive resources to make good decisions. Nevertheless, there are many occasions when we can consciously deliberate and decide on one course of action over another.

Deliberating carefully is a lot like conducting an inner argument with oneself. As we deliberate, we are moved to act in one way or another, at least in part through the consideration of reasons. There are different levels of responsibility associated with different kinds of decision making. Unless we are in a position of great responsibility, many of our decisions concern matters that are not very important to others. Examples of decisions where the stakes are low and we can afford to be sloppy might include:

> What color shirt should I wear?
> What should I watch on TV?
> Should I eat an apple or a doughnut?

> When a decision concerns an important matter, it is vital that we eliminate as many distractions as possible in order to *carefully consider evidence* and *evaluate arguments* for competing courses of action.

These are private matters and for the most part, when the well-being of others is not directly affected, being irrational is harmless—although keep in mind, those doughnuts can add up. However, while it is often perfectly fine to be irrational, almost all of us will face decisions where the stakes involved are high and where it is morally required that we reason carefully. This is especially true when our decisions have direct consequences for other people.

Even when it comes to decisions that only directly affect the decision maker personally, there are different levels of importance. Consider a student who must decide whether to enroll for another semester of college. This is a relatively important decision, but it principally concerns the student's own well-being and therefore the decision is not as morally significant as decisions concerning the well-being of others might be.

Before we examine how a student might reason about this decision, it is worth pausing and reflecting on the factors that a student *should* consider while making this decision. While there are many things that a student should consider, in practice any individual student is unlikely to consider all of them. Under the pressure of having to make an important decision, we may be more likely to reason in an incomplete and sloppy manner. In any event, let's imagine that the student's deliberation goes something like this:

"The deadline for registration is approaching. Should I enroll for the spring semester?

I'm not sure. Look at Mark Zuckerberg and Bill Gates, they never finished their degrees and they did just fine. In fact, people think it's kind of cool that they never finished college. I hated that economics professor from last semester, she was so mean to me. I can always go back to my old job at that store. It wasn't so bad. Maybe I could complete my degree in my spare time at Firebird University Online. I love watching videos online. College is filled with pointless work. When will I ever use what I learn in that sociology class anyway? Another semester in college would be an exhausting and expensive waste of my time. It's just not worth it. Anyway, someone posted an article the other day on my Friendface timeline saying that college is obsolete. . . .

But on the other hand, I don't want people to think I'm a quitter. I like learning new things and was bored out of my mind working in that store. Working there was even more boring than that formal epistemology course with that Irish professor. All my friends are going to finish, so I guess I should too. Don't people who have a college degree make more money over the course of their lifetimes? I'd hate to be poor, so maybe I should stick with it. It's a lot of work, but I'm not lazy. My parents would be so disappointed if I dropped out. I'd be a dropout! And think of all the work I've already put into this thing…

OK, I've decided. . .it's better for me to stay in college so I'll enroll for the fall semester."

This fictional example contains at least one very unrealistic feature: Chances are, when we really think problems through for ourselves, the reasons for and against some course of action will not be as neatly distinguished as in our example. Instead, before we begin thinking carefully, these factors pro and con will be jumbled together in our thinking. In fact, notice that by separating the student's reasons for and against enrolling for another semester, we are already formalizing their reasoning to some extent.

Sometimes people facing important decisions will organize their thinking by creating columns for reasons pro and con. This is a very simple technique for formalizing our decision making. You may have already used this technique in your own decision making at some point. Notice the effect it has on our deliberation. To begin with, it certainly helps us to slow down and to think though our decision a little more carefully than we otherwise might. Listing pros and cons also helps us to include additional considerations. In the fictional case above, you can probably already see that the student has not considered all the relevant reasons for and against the course of action.

> Separating the reasons for and against some course of action is a very simple example of how we can begin to formalize our reasoning.

The example of decision making given here is meant to be sloppy but there is some reasoning here nonetheless. At this point, common sense allows you to tell that some of the reasons this student entertains are good, some are not. As you follow this student's reasoning, you might be tempted to help out, by pointing out the strengths and weaknesses of the reasons under consideration. At the very least, some obvious questions have probably occurred to you.

Let's begin by thinking about the evidence that figures in the reasoning. We should begin by asking the very general question:

> What evidence supports the factual claims that are introduced into the argument? Are the sources of this evidence reliable? For example: Is it really true that graduates earn more than noncollege graduates? Where did the student get this information?

Beyond asking about sources and evidence, we should also be critical in our assessment of the way that the evidence is talked about. To begin with, we should think about the precision of the terms involved in the debate. Often, claims are introduced into arguments whose precise meaning is unclear. Since we hear certain words frequently, we are sometimes lulled into thinking that we know what those words mean. When the meaning of the terms involved are unclear or are introduced uncritically, they can have significant effects on the course of an argument. Introducing unintended prejudices or concealing unexamined agendas. For example, we could ask:

> Is higher education really *obsolete*? What precisely would it mean to make a claim like that? Who determines what is meant by *obsolete*?

Even in cases where the evidence is true and stated precisely, we should consider whether it is relevant to the problem under consideration.

> Are Mark Zuckerberg and Bill Gates representative cases to focus on when we consider the fates of people who do not complete their degrees? Among the very large number of people who do not complete their degrees some are almost certainly going to be very successful. But what happens to the average member of that group?

> Is the student's experience with the economics professor really relevant to the decision?

Some of these questions concern specific facts or pieces of evidence, some concern the methods by which the student acquired their evidence, and some concern the weight that certain pieces of evidence have in the student's decision making. Even if you agree that the student made the correct decision, a number of questions like this are likely to occur to you. Furthermore, you might also be able to pick up on a number of problems in the student's reasoning that have nothing to do with evidence or the facts per se. To put it another way, in addition to being concerned the premises that this student brings to their deliberation, we should also be concerned about the inferences that led the student to conclude as they did. Part of what it means to reason badly is to move from one thought to another in illegitimate ways. Some of these errors may already be obvious to you; others will become obvious to you as you continue your study of logic and critical thinking. You probably noticed that when the student says

> "All my friends are going to finish. So I guess I should too"

they are committing a pretty simple error in reasoning.

To take another, slightly less obvious error in the student's reasoning; it is not true that we ought to continue on some course of action simply because we have already given it our time or other resources in the past—as we shall see, this is a common mistake in reasoning known as the *sunk-cost fallacy*. We will examine this common fallacy in detail along with others later in this book.

Clearly all of us would prefer to make our decisions for good reasons rather than bad. The decision as to whether to continue one's education is important enough to deserve some careful thought.

> *Making good decisions involves doing our best to reason well. Good decision making requires careful deliberation with respect to the arguments for and against some course of action.*
>
> *The first step is to learn how to judge the quality of arguments.*

2.3 How to Begin Evaluating Arguments

In order to carefully evaluate the diverse range of informal arguments we encounter, it helps to adopt a generous attitude. Almost all the arguments one encounters in daily life are messy. However, just because an argument is not presented in the style of a legal brief or an article in an academic journal does not mean that it should be dismissed. We are interested in the pursuit of truth and we should be open to the truth coming in unexpected, subtle, and sometimes messy packaging.

The first step is to approach important arguments assuming that our conversational partner or the author of the text is actually making an argument, has reasons for their view, and has arrived at that view in a rational manner. These assumptions might be incorrect, but in important matters, it is generally worth the risk of being too charitable. The philosopher Donald Davidson called this *the principle of rational accommodation* by which he meant that one ought to try to interpret the statements of others in a way that maximizes our points of agreement and makes as much sense of what others are saying as possible. When one hears an argument from someone who holds views that are at odds with one's own, the first step is to find a sufficient number of points of agreement to allow the conversation to start fruitfully. One might later find out that this initial generous reading was misguided, but as a general rule when one reads or listens

> Remember, our goal is not to *win* a debating contest with the people we talk to or with the authors whose work we read. Instead, our goal is to find the best arguments we can and to make the best decisions we can.

to others, one should try to give their arguments the most charitable interpretation possible.

In most ordinary contexts, it is necessary to listen and read carefully and charitably in order to be in a position to even identify the parts of an argument. As you browse the Internet or consume other kinds of media, you will encounter many examples of people seeming to insist, seduce, sell, entertain, or distract. It is often the case that even locating the logically relevant parts of arguments in this context will be a challenge. But even many of the most trivial and silly seeming pieces of media can be understood as presenting arguments.

Obviously, it would be impossible to say determine whether an argument is good if we were unable to identify its main contention. As mentioned above, the main components of an argument are the following:

> **The main contention or conclusion**: The claim that the argument as a whole is intended to support or justify.
>
> **The assumptions or premises:** The claims that serve as common starting points for the person making the argument and his or her audience.
>
> **Inferences:** The moves that connect the assumptions or premises to the conclusion.

In ordinary experience, identifying the parts of an argument can often be quite challenging. The source of this difficulty is the diverse, multi-level, and highly contextual character of human discourse. Clear, well-written arguments, of the kind you can find in good philosophical or legal writing are not what we regularly encounter on social media, in the mass media, in politics, or in commercial life. To think critically about the kinds of content we receive from these sources, the first step is to take the sources seriously. Assuming that the creators of the material we consume are serious people with something that they sincerely wish to convey to us is, for the most part, to assume something false. Nevertheless, treating them as though they were making an argument, and as though they hoped to rationally persuade us of some contention, is a good strategy for avoiding the seductive manipulation of most contemporary cultural products.

No matter what the source, we must first judge the relevance of the individual sentences we find in an argument. We will soon realize that in most contexts; in opinion pieces, conversations, blog posts, and editorials, **most of what is said or written is not directly relevant to the argument**. An important part of our task is to distinguish those sentences that do not contribute to the argument from those that contribute directly to the real structure of the argument.

As we approach a piece of text or a cultural product of some other kind, we will begin by asking the following questions:

Can this be read as an argument?
If so, what is the purpose of this argument?
What is being assumed?
Are there any unstated assumptions in this argument?
How does the argument move from the assumptions to the main contention?

To decide whether what we are listening to or reading actually is an argument, we must determine whether it is a finite sequence of sentences that is intended to convince us of some contention or conclusion. Once we have decided that we are dealing with an argument, we can begin assessing whether it is a good or a bad argument.

What is the purpose of this argument?

The first step in evaluating an argument is to find its main contention or conclusion. The conclusion is, of course, what the speaker or author is attempting to convince his or her audience to accept. When we are looking for the main contention, we are looking to understand the overall point of the written or spoken argument. In many informal contexts, a speaker or an author might not be clear about his or her conclusion or might be arguing for multiple (or even mutually incompatible!) conclusions. When we are not being careful with our talk and writing, it is easy to lose track of what we are arguing for.

There might not be a point

It is also possible that someone appears to be making an argument, but there is no conclusion to be found. As you read or listen to an argument, it is usually worth assuming that there is a conclusion, even if the conclusion is not stated in the argument. As mentioned above, this requires us to become charitable readers and listeners. We will discuss the role of charity in the interpretation of arguments below.

Finding the conclusion

In more formal contexts there are a set of tell-tale markers that indicate the presence of a conclusion. These include the presence of words and terms like:

> ***therefore, hence, because of this, accordingly, thus, this entails, this suggests, this implies, this proves, I will argue that, my thesis is . . .***

Phrases or words like these indicate that the author means to mark some kind of conclusion. However, these words might appear repeatedly in a single text or conversation and this is sometimes due to the presence of subarguments for subconclusions that support the main contention of the argument. It might be the case that in order to establish his or her main contention, an author must make a number of subarguments along the way. These subarguments establish claims that in turn are intended to support the main contention.

As we are looking for the overall purpose of the argument, its main contention, it is important to note that the main contention can appear anywhere in the course of an argument. It does not necessarily appear at the beginning or the end of a text or conversation. It is also important to avoid prematurely judging an author's or a speaker's main contention. Frequently, we can misidentify some subconclusion or some premise as the main contention of an argument if we are in too much of a hurry. With complex arguments it is worth reading or listening carefully and patiently in order to determine precisely what the point of the entire argument actually is.

Once you determine the purpose of an argument, the next step is to ask whether each sentence contributes to the argument and how it contributes to the argument for the main contention. At this stage, we can keep the following basic checklist in mind for each of the sentences.

Initial Argument Analysis Checklist

- Is the sentence a premise?
    - Does it introduce factual evidence or mention some shared assumption?
- Is the sentence responding to an objection?
    - Is the author attempting to respond to a possible counter-argument?
- Is the sentence derived from some previous sentence via some piece of reasoning?
    - Is the author making an inference of some sort? Can you tell how the author made the inference?
- Is the sentence stating a conclusion?
    - Perhaps the point of the argument is never explicitly stated anywhere in the argument.

2.4 Subarguments

Many arguments will contain subarguments or nested arguments in support of the main contention of an argument. In this section, we will explain the role of subarguments and how we can map them as we find them in a complex argument. The reason that an argument might have subarguments is pretty straightforward. The main contention of a complex argument might require the proponent to establish or defend some nonobvious premises for the main contention. These might require their own subarguments.

Suppose that you are arguing for the following contention:

Students should be given government subsidized loans to support their education

> Let's assume that this is the main contention of your argument.

Perhaps you don't agree with this contention, but suppose for the sake of our work here that you do. How would you go about rationally convincing someone to accept your contention? At this point, you know that you would begin with some shared set of premises and move using legitimate steps to the conclusion. Anyone who accepts the truth of your premises and can follow the inferences of your argument is obliged to accept your conclusion. However, how your argument goes depends on the premises that you can take as your starting point. Audiences differ with respect to which premises they will accept and which will need to be justified. It is not always a simple matter to know where to begin. It might be necessary to provide an argument to convince your audience to accept as true, premises that you think are completely obvious.

For example, we can imagine someone objecting that subsidized loans are not the most efficient way of supporting education. Another audience might accept that loans are an efficient means of giving financial assistance to students. Another audience might not be concerned with economic efficiency in cases like this at all. Thus, where you start and which subarguments you will need to introduce is partly determined by the needs and commitments of your audience.

Let's imagine a fiscally conservative or perhaps a libertarian audience who question the very idea of being compelled by the government to pay for good things for strangers. Such an audience might wonder why the government should force citizens to use their resources to support the education of others. In this case, in order to argue for the main contention, it would first be necessary to convince such an audience that:

*Governments **should** use resources to financially support education.*

> You need to argue for this sub-conclusion in order to convince this fiscally conservative audience of your main contention.

Notice that the objection we are considering here is not concerned with efficiency, but rather with the more fundamental principle that we should collectively support education. In order to respond to this objection, the proponent of the main contention might need to defend the idea that:

Education is a public good.

> In order to convince the fiscally conservative audience to accept that the government should support education, you need to establish that this support is not simply a private benefit for the student, but is a public good, like defense, law enforcement, or clean air.

By arguing that education is a public rather than solely a private good, the proponent would be arguing that when a young person is educated, it benefits all of us; it is not just a benefit for the student. But even if the proponent were able to establish this, he or she might still need to argue that:

Governments should expend resources in pursuit of public goods.

> Even if the audience accepts that education is a public good it might be still necessary to defend this basic principle.

Clearly, that the kind of subarguments that you need to make to convince your audience is dependent on the prior beliefs that your audience brings to the discussion. A politically left-leaning audience is likely to simply accept that governments should financially support education. If this is the case, then there will be no need to make the additional subarguments concerning public goods. However, it might be the case that other subarguments are necessary.

Consider an audience of politically progressive economists who doubted the efficiency of the subsidized loan system. An audience of this kind would need a very different line of argument in order to be convinced of the main thesis. They might accept that governments should support public education while doubting that loan subsidies are a good way of offering this support.

Consider another audience who believes that all students are entitled to free access to higher education. This audience would resist the idea that students should be burdened with any financial cost for their education whatsoever, including subsidized loans. This audience will need a very different kind of argument again from the other two audiences that we have considered.

In ordinary contexts, the structure of subarguments depends in part on the presuppositions that our audience brings to the argument. In the two examples above, consider how you would argue for the contention. What subarguments would you need to introduce? What premises do you need to defend in order to support your contention?

Once you have thought about what you would need to do in order to shape an argument for a particular audience, you should start thinking about how arguments that you encounter from others are structured. As you begin to conduct your analysis, you can ask the following questions?

Who is the target audience for this argument?

Here you can consider the venue where you encounter this argument. Is it directed at readers of a partisan blog, a scientific journal, or a respected newspaper?

What background knowledge, presuppositions or assumptions would this audience bring to their reading?

Audiences will vary with respect to what they know and what they value. It is useful to keep that in mind in order to help you to consider alternative perspectives and alternative kinds of audiences.

Imagine a different kind of audience. Which of these presuppositions would they reject or at least challenge?

Sometimes, arguments that are directed toward particular audiences suffer from an echo chamber effect, with certain beliefs or arguments being repeated or even amplified without being questioned. The reason for this effect is that people who are like-minded tend to reinforce one another's beliefs. The mere fact that others agree with my views does not guarantee that my views are correct. Nevertheless, if I am never questioned by my audience, I can easily gain a false sense of confidence with respect to those beliefs.

Can the argument be reconstructed in a way that would convince this imagined alternative audience?

It is worth considering an argument from the perspective of an audience that does not share the assumptions or presuppositions of the target audience. Is this argument designed solely to convince a particular kind of audience, or does it have more universal standing?

2.5 An Argument about Death

Philosophers have provided arguments concerning the deepest and most important topics in human life. Unlike mystics or poets, the main business of philosophers is to give good arguments for their claims. This partly explains why philosophers have studied the nature of argument so carefully.

Philosophy is concerned with the most basic questions about reality, knowledge, existence, and goodness. Over the millennia, philosophers have offered arguments in support of their claims and have attempted to rationally persuade others of the correctness of their view. For example, consider your own death. Unless you are a very precocious young reader, it is a pretty safe bet that you, the reader of this book will be dead within 80 years. How should we react to the unsettling thought of our own nonbeing? Should we be afraid to die?

According to the philosopher Epicurus (341–270 BCE), it is a mistake to fear death. Epicurus hopes to convince his readers that we should not fear death and he has a famous argument in support of this claim.

Epicurus provides an argument that is intended to convince us by appealing to our rational capacity. By contrast, he is not asking his readers to accept his conclusion on the basis of faith, or because it pleases, flatters, or comforts us in some way. Religious authorities, for example, often insist that you accept statements on faith; without a rationally satisfying explanation. They simply ask that you take their word for it. Similarly, an artist might provide us with an attractive vision of a life without the fear of death in a novel, a poem, or a movie. We might decide that a life without a fear of death is esthetically pleasing and might therefore opt to abandon our fear of death. There is some value to religion and the arts, but Epicurus and other philosophers are engaged in a very different enterprise. His task is to provide a convincing justification for thinking that one should abandon one's fear of death that does not appeal to faith, emotion, or esthetics. In so doing, he is attempting to rationally persuade us, rather than influencing us in nonrational ways. His argument was presented in a letter to his friend Menoeceus, it goes roughly as follows:

When we are dead, we no longer exist.	***(First Premise)***
Only existing things can be harmed.	***(Second Premise)***
We cannot be harmed by something that has not yet happened.	***(Third Premise)***
We are either alive or dead.	***(Fourth Premise)***
We should not fear harmless things	***(Fifth Premise)***
We are not harmed by death when we are dead because we do not exist. ***(Inference from the first and second premises)***	***(Inference)***
We are not harmed by death when we are alive because we are not dead yet. ***(Inference from the first and third premises)***	***(Inference)***
We are either alive or dead ***(the fourth premise)*** *and we are not harmed by death when we are alive and we are not harmed*	

by death when we are dead. ***(as we showed via the previous two inferences)***
Therefore we are never harmed by death ***(Inference)***

Given that we should not fear harmless things (fifth premise)
and since death is harmless, we should not fear death. ***(Inference)***

We should not fear death ***(Conclusion)***

How would you judge this argument? Do you accept its conclusion? In order to rationally reject the conclusion you must

deny its initial assumptions (show that the premises are false)

or

show an error in the reasoning from the premises to the conclusion (show that the inferences are bad).

Perhaps you think that death is something to be feared insofar as it deprives you of good things that you might have had were you to live long enough. Perhaps you deny that death is the same as non-existence. Perhaps you think that you can be harmed, even if you do not exist. In each of these cases, you would be differing with the author concerning one of the premises of the argument.

If you do not accept his conclusion and wish to consider yourself rational, then you must find fault with either the premises or the inferences. If you accept his premises and you accept the quality of his reasoning, then you must accept his conclusion. Doing so means that you believe that you have no rational basis for fearing death.

Notice that rationally accepting or rejecting his argument is not a matter of liking or disliking the conclusion. Indeed, one might rationally accept the argument and yet still continue to feel fear at the thought of one's death. To a certain extent, our feelings are not directly amenable to rational persuasion.

Our feelings are important, but they are not directly relevant to the evaluation of arguments. Instead, as we are beginning to see, deciding whether to accept any argument depends solely on evaluating:

The truth of the premises

and

The correctness of the inferences

It is difficult to think clearly about a topic like death because of the powerful emotional factors at play. Notice that in this argument, Epicurus is attempting to use rational persuasion to change our view of one of the most emotionally charged topics there is; our own death. As we shall see, one of the salutary effects of formal reasoning is that it helps us to abstract away from the distorting effects of feelings, biases, and personal preferences in order to evaluate the merits of an argument.

Unfortunately, we tend to be confused about the relationship between our feelings and our commitment to the truth and falsity of claims. We tend to accept beliefs that make us feel comfortable and to reject those that make us uncomfortable simply because of our feelings of comfort or discomfort. This is a regrettable tendency. Being a good thinker means being able to evaluate an argument independently of whether we like or dislike its conclusion. Very often, the truth is unpleasant.

Many of us have had the experience of complex and emotionally charged problems that nevertheless required us to think clearly. As an adult, it is difficult to completely avoid painful decisions. When we are faced with choices concerning health, pregnancy, relationships, career, and finance, for example, it is nearly impossible to eliminate the influence of emotion. However, decisions like these are often those in which careful deliberation and rational evaluation of arguments are most important. **Often, the most emotionally charged decisions are those where the cost of irrational or arbitrary thinking is the highest.**

Emotions as Warning Signs:

Whenever one feels emotionally disturbed by a decision; when one feels excited, proud, angry, frustrated, or defensive, one is liable to make errors in reasoning. Given that one cares about the decision, these occasions merit extra care.

Not all arguments will be as systematic as our presentation of Epicurus' argument. In fact, most arguments (even arguments given by philosophers) are informal and they generally require a charitable interpretation and careful reading. Frequently, it will take some work for us to figure out what the main contention of the argument actually is, what assumptions are being introduced, and what inferences the argument contains. For example, in my presentation of Epicurus' premises, I am including assumptions that are not explicitly stated in the passage from the letter to Menoeceus. Unpacking the parts of the argument precedes the task of evaluating its inferences and premises. In the next chapter, we will examine informal arguments in more detail.

Exercises for Chapter 2

1. Consider an important decision that you have had to make that required you to evaluate the pros and cons of different courses of action. How did you compare the pros and cons? How much weight did you give to them? Were they always comparable? For example, when comparing moral versus financial considerations how does one weigh one against the other?

2. Consider occasions where the principle of rational accommodation becomes difficult to apply? What are some characteristic features of such occasions?

3. Find another argumentative essay or editorial online. Consider how the choice of audience determines the characteristics of the argument. How would the argument be different if the audience were different? Here, think specifically of the kinds of subarguments and hidden assumptions that the argument contains. Which assumptions would be foregrounded and how would the subarguments differ given the new audience that you imagine.

4. Consider an emotionally challenging decision from your own life. What role did emotions play in the decision? Can you imagine a scenario in which strong emotions might help you to make a good decision? What about a scenario in which someone felt no emotions whatsoever? Would that person even be capable of making a decision?

3 Informal and Dynamic Arguments

In this chapter, we tackle the challenging domain of informal arguments in more detail. Most of the arguments that one encounters in ordinary life are not presented formally or clearly. With Chapter 2, we have taken the first step in approaching arguments; we have seen how to begin analyzing an argument in order to determine its parts. However, arguments not only contain conclusions, inferences, and premises, they often also contain sub-arguments. What is a sub-argument? In addition to having a single line of argument with a single conclusion, it is common for complex informal arguments to consist of a number of sub-arguments that contribute to the overall contention of the larger argument. The purpose of the argument as a whole is to support the central conclusion or the main contention of the argument. However, along the way, it might be necessary to present some number of arguments to establish one or another step or part of the overall argument. Mapping these sub-arguments and untangling the overall structure of a large complex argument is the purpose of this chapter.

In addition to arguments having sub-arguments, it will often be the case that they have a dynamic form. By saying that arguments have a dynamic form, what is meant is that they will be constructed in real time, in response to the changing conversational context or they will be revised in light of responses from an audience.

Unpacking the logical structure of real arguments in informal contexts requires a great deal of careful and sensitive reading and listening. One needs to be sensitive to their complicated sub-argument structure and to the complexity of their dynamic form. One also needs to become a generous listener and reader. A critical thinker who is both sensitive to complexity and approaches arguments charitably can find the truth in unexpected packages.

3.1 Mapping Sub-Arguments

As we read complex arguments, it helps to begin by listing the main points of the text as a way of distinguishing the components of the argument. Figuring out the main points of the text is equivalent to getting the gist of the argument. What is the overall point of the argument and what are some of its landmarks. Once we establish these general characteristics, we can begin to map the structure of the complicated argument in order to evaluate the formal properties of the argument more fully.

Recall our initial argument analysis checklist presented in Section 2.3. When we approach an actual argument, say, for example, an editorial in a newspaper, we can use our checklist to label each of the sentences as best we can.

- Is the sentence a premise?
    - Does it introduce factual evidence or mention some shared assumption?

      If so LABEL it *PREM*

- Is the sentence responding to an objection?
    - Is the author attempting to respond to a possible counterargument?

      If so LABEL it *OBJ*

- Is the sentence derived from some previous sentence via some piece of reasoning?
    - Is the author making an inference of some sort? Can you tell how the author made the inference?

      If so LABEL it *INF*

- Is the sentence stating a conclusion?
  - Perhaps the point of the argument is never explicitly stated anywhere in the argument.

    If so LABEL it *CONCL*
- Is the sentence irrelevant to the argument?

    If so LABEL it *IRR*

Every sentence should be labeled.

Consider the case of a discussant who is arguing for a free-trade agreement between two countries. The free-trade agreement would end tariffs on imported goods, but it would also lead to unemployment in industries that were previously protected by barriers to trade. Arguments in favor of, or against free-trade agreements, are complex and involve a range of factual and theoretical claims in economics that might be difficult for us to verify.

Expensive Cotton Brought to you by the Madman of Fanta Sé

The people of Happyland have lived with the threat of war for many generations now. We must do whatever we can to avoid the costly and damaging conflicts that have plagued our citizens in the past. Readers will recall the devastation of the 10-year war against the Gnomes that left our northern provinces in the grip of severe economic problems for far too long. As we all know, countries that trade together actively are unlikely to go to war and we believe that increased economic activity between our country and our neighbors to the south in Resourcia is a critical part of ensuring a continued and sustainable peace. We fought the gnomes in defense of our principles. Happyland is a land of liberty and it runs counter to our basic values for the government to interfere with our individual right to trade with whomever we wish. Let's face reality and recognize that the border is a fiction; a mere line on the map. It is morally unconscionable for our federal government to use this meaningless line as a way of keeping our businesspeople from finding the most efficient and economically effective trading arrangements available. Recent statements by the Governor of the State of New Mystery Mr. Zebedee Fancyperm stating that he wants to reinforce the border with additional chicken-wire and duct tape should be treated with the derision that they deserve. None of us should fall victim to his insane pronouncements against international cooperation. Governor Fancyperm is a madman who fails to understand the basic economic benefits of free-trade. His objections to the proposed agreement with Resourcia should be ignored as ill-informed demagoguery. He has misunderstood the moral and economic arguments for the agreement and should return to his governor's mansion in Fanta Sé in shame. This is a federal and not a state matter. Governor Fancyperm is an embarrassment to the state of NewMystery and he has no business meddling with the economic future of our nation. Economists all agree that the proposed free-trade agreement will remove artificial trade barriers, leading to greater economic efficiency, increases in trade and significant growth in the economy of Happyland. Because of taxes on imports of cotton from Resourcia, we live with the ridiculous circumstance wherein the farmers of Happyland irrigate their cotton fields with our precious and limited water supply in order to bring artificially expensive cotton to market. There is an abundance of cheap cotton in Resourcia and even though some of Happyland's cotton farmers will suffer in the short-term, the vast majority of Happyland's citizens will benefit from access to cheap cotton. Let us remember all those who have died in the war against the gnomes and should pay no heed to the madman in Fanta Sé

In order to make his or her case, the author might first need to convince readers that free trade is something that we should favor. As we saw in the previous chapter, not all audiences will require an argument in order to be convinced of this claim. Most modern economists, for instance, favor international free trade. However, if it is necessary to make the case for free trade, then the argument is likely to include sub-arguments of various kinds, containing responses to possible objections, etc. Notice that the set of sub-arguments that one finds in an argument is determined by the proponent's view of their audience.

Let's read the argument and apply the initial checklist to each sentence before we begin outlining the sub-arguments. Notice that there will be some sentences whose role in the argument is uncertain. Is the sentence a premise, or is it irrelevant rhetoric or bluster? Is the sentence the result of an inference or is it simply an assertion? Often it will be possible for reasonable people to legitimately disagree on the interpretation of a sentence's role in an argument. Let's apply the checklist to each sentence in the argument. Perhaps you will disagree with the labels I have chosen.

Expensive Cotton Brought to You by the Madman of Fanta Sé

The people of Happyland have lived with the threat of war for many generations now. PREM

We must do whatever we can to avoid the costly and damaging conflicts that have plagued our citizens in the past. PREM

Readers will recall the devastation of the 10-year war against the Gnomes that left our northern provinces in the grip of severe economic problems for far too long. PREM

As we all know, countries that trade together actively are unlikely to go to war and we believe that increased economic activity between our country and our neighbors to the south in Resourcia is a critical part of ensuring a continued and sustainable peace. PREM

We fought the gnomes in defense of our principles. IRR

Happyland is a land of liberty and it runs counter to our basic values for the government to interfere with our individual right to trade with whomever we wish. PREM

Let's face reality and recognize that the border is a fiction; a mere line on the map. PREM

It is morally unconscionable for our federal government to use this meaningless line as a way of keeping our businesspeople from finding the most efficient and economically effective trading arrangements available. CONCL (?)

Recent statements by the Governor of the State of New Mystery Mr. Zebedee Fancyperm stating that he wants to reinforce the border with additional chicken wire and duct tape should be treated with the derision that they deserve. IRR (?)

None of us should fall victim to his insane pronouncements against international cooperation. CONCL (?)

Governor Fancyperm is a madman who fails to understand the basic economic benefits of free trade. IRR

His objections to the proposed agreement with Resourcia should be ignored as ill-informed demagoguery. IRR

He has misunderstood the moral and economic arguments for the agreement and should return to his governor's mansion in Fanta Sé in shame. IRR

This is a federal and not a state matter. PREM

Governor Fancyperm is an embarrassment to the state of New Mystery and he has no business meddling with the economic future of our nation. IRR

Economists all agree that the proposed free-trade agreement will remove artificial trade barriers, leading to greater economic efficiency, and increase in trade and significant growth in the economy of Happyland. PREM

Because of taxes on imports of cotton from Resourcia, we live with the ridiculous circumstance wherein the farmers of Happyland irrigate their cotton fields with our precious and limited water supply in order to bring artificially expensive cotton to market. INF

There is an abundance of cheap cotton in Resourcia and even though some of Happyland's cotton farmers will suffer in the short-term, the vast majority of Happyland's citizens will benefit from access to cheap cotton. PREM

Let us remember all those who have died in the war against the gnomes and should pay no heed to the madman in Fanta Sé IRR

Having read through the argument in its entirety, we began by getting a sense for the overall purpose, or the gist of the piece. What would this author want me to think or do as a result of reading this piece? What purpose motivates the author to write this piece?

It is important to read the entire piece before assuming too quickly that we understand the point of the argument. Now that you have examined the argument carefully, you have probably identified its main contention:

> *Main Contention: Happyland should sign a free-trade agreement with Resourcia.*

Notice that as we determine the overall point of the article, we may end up discovering that most of what is written is irrelevant to the argument for the main contention. For example, in the present argument, the insults directed toward Governor Fancyperm and the repeated mention of the war against the gnomes has relatively little to do with the main contention. Because so much of the arguments we read in newspaper editorials or on blogs have relatively little to do with the actual argument, it is easy for us to incorrectly interpret the purpose of this kind of writing.

Why, for example, don't we regard

> "We must do whatever we can to avoid the costly and damaging conflicts that have plagued our citizens in the past."

as the main contention of the editorial?

While it bears some resemblance to a contention, there are at least two reasons why it would be a mistake to interpret this as the main contention of the piece.

1. The claim that we should avoid costly and damaging conflicts is relatively uncontroversial. It is difficult to imagine anyone feeling the need to present an argument in support of this claim. Instead, it is introduced early in the argument as an assumption in the service of the overall argument.
2. There is no attempt made in the later parts of the argument to convince someone who is in favor of damaging conflicts that they should change their view. By contrast, the target

> audience of the piece is a reader who might disagree with, or be unconvinced by the actual main contention; the claim that Happyland should sign a free-trade agreement.

The editorial as a whole is directed toward encouraging readers to favor a free-trade agreement with Resourcia and there are three significant lines of argument presented in its support. As we follow the main points of the argument, we find that it has a structure like the following:

Happyland should sign a free-trade agreement with Resourcia
[This is the main contention of the argument]

Why?

If we are engaged in productive trade with one another, we are less likely to go to war with one another

Why?

We have a right to trade with whomever we want and national governments should not stand in our way.

Why?

Because it will provide economic benefits for Happyland

Why?

Because there will be increased trade between Happyland and Resourcia

Why?

Because removing trade barriers encourages economic activity

Why?

Because markets will force each country to specialize in those goods and services it can most efficiently produce.

OBJECTION: Happyland's cotton farmers will be harmed by lifting tariffs on imported cotton.

RESPONSE: While it's true that cotton farmers in Happyland will suffer; the vast majority of Happylanders will benefit from access to cheap imported cotton. This benefit outweighs any harm to the farmers.

At this point, we can consider that the three separate lines of reasoning are provided in support of signing a free-trade agreement with Resourcia.

(a) If we are engaged in productive trade with one another, we are less likely to go to war.
(b) We have a right to trade with whomever we want and national governments should not stand in our way.
(c) The agreement will provide economic benefits for Happyland.

Imagine a discussion wherein the audience member accepted the first and second reasons, but did not accept the third. In this case, the proponent of the thesis would have to provide reasons to support (c). Each of these reasons, in turn might need to be justified or defended depending on the audience in question. In this case, there is an extended sub-argument for (c) but none for (a) and (b). In an informal setting, deciding which reasons need additional support will be a matter of determining what your audience will accept.

As we begin to read argumentative essays, we can practice locating precisely those elements that play a role in supporting the conclusion and responding to objections. At this stage, our goal is simply to identify the parts of a text that are actually doing some argumentative work. We are not yet evaluating the quality of the arguments. This will follow shortly.

As we study the structure of the argument, it is worth noting the points at which an author considers possible objections from an opponent. In this case, the author considers only one objection, albeit an important one, namely that free trade in cotton will harm the farmers of Happyland. The author responds to this objection by claiming that this harm will be outweighed by the benefit that cheap cotton will bring to the majority of Happyland citizens. We should reflect on whether this is a good response to the objection.

A critical reader will consider whether relevant objections have been considered at each important step in the argument.

3.2 Uncovering Assumptions, Reading Charitably, and Reading Critically

Most arguments in ordinary experience do not explicitly state all the relevant premises or assumptions that serve as the basis for their conclusion. In the argument we have just considered, the author assumed, for example,

- that peace and prosperity are desirable,
- that economics is a discipline that can reliably guide public policy,
- that the economic benefit to the majority of Happyland citizens outweighs the harm to its cotton farmers, and
- that we have a moral right to trade with whomever we wish

You may not accept all of these assumptions. For example, there are good reasons to hesitate before embracing the idea that one should be allowed to trade with whomever one wishes, for example. Surely, we can imagine conditions where it is appropriate for governments to restrict trade with enemy nations, terrorists, with small children, with people in prison, etc.

If you know your audience reasonably well, you are likely to know what information you can assume as a shared starting point. For the sake of efficiency, when we are attempting to convince our audience of some claim, we usually focus our argument on the points of disagreement, or contention. When an author is unsure of their audience's beliefs, more care and explicitness is required.

The assumptions that support an argument are indicated by using words or terms such as

I assume that. . .

Given that. . .

Because of. . .

Since. . .

The reasons for this are. . .

Insofar as. . .

To find the **hidden premises** in an argument, we must first develop a **charitable interpretation** of the proponent's argument.

> **A charitable interpretation** is one that assumes that the author or speaker:
>
> aims at the truth or at the very least is not being deliberately deceptive.
>
> is rational; that they are not willingly asserting incoherent or contradictory claims.
>
> has a store of background knowledge or contextual knowledge that the author assumes is shared by readers or listeners.

As charitable interpreters, we should understand the author of an argument to be working with a particular understanding of the competence and expertise of their audience. For example, if the author is addressing a group of astronomers they can take it for granted that their audience will have more background knowledge concerning, for example, the nature of quasars than an audience of nonexperts. An argument concerning the nature of quasars, presented to a group of astronomers, will contain many more unstated assumptions than we would find in, for example, an introductory astronomy textbook or in a news magazine.

Thus, a charitable reconstruction of an author or speaker's arguments involves attempting to figure out and include their relevant unstated premises. We should also aim to give the strongest plausible interpretation of the role of those premises in the argument that attempts to honor the intentions of the author or speaker as best we can determine them.

By providing a charitable interpretation of a text, we may end up providing the author with a stronger argument than they originally presented. This should not worry us too much. After all, our purpose in critically reading and listening is not to defeat or outsmart the author or speaker in a debating competition. Debates are interesting and fun, but our goal is not to demonstrate our superiority over our opponents. Instead, we are interested in understanding whether or not to accept the conclusion of the argument independently of any competitive considerations.

It is important to avoid personalizing our considerations of arguments. It might be the case that an author holds some belief for irrational or for some otherwise objectionable reason. Whether you accept some, conclusion should be determined by the quality of the argument, not the history or the character of the speaker or author.

As we consider an argument, our task is to decide whether or not to accept the conclusion. As we have seen, the first step is to sketch out the parts of the argument to the best of our ability. With a charitable presentation of the argument in hand, we will be in a position to begin asking some questions concerning its quality.

1. What evidence could be mentioned in support of the main contention?
2. How reliable is this evidence?
3. Is this evidence relevant to the argument for the contention?
4. Are the inferences that are made in the course of connecting the evidence and the main contention legitimate?
5. Does the argument successfully address the relevant objections?

In later chapters, we will explore ways to evaluate the legitimacy of inferences in more detail. At this point, our goal is simply to begin to uncover the reasoning that takes an argument from premises to conclusions. We will soon be in a position to evaluate the inferences at the heart of a proponents reasoning.

The trouble with ordinary arguments is that the inferences are often difficult to track. If it is difficult to see how an author arrives at his or her conclusion from the text alone, we are forced to charitably reconstruct their reasoning as charitably as possible.

As we have seen, an inference is simply a move from one claim about the way things are (or the way things could be) to another. Much of our daily life is spent making some kind of inquiry or engaging in some kind of reasoning. As we draw conclusions from evidence or assumptions, we are sometimes misled by habits of thought, strong associations, or powerful emotions. If our reasoning is faulty, we can be led to mistaken beliefs. Basing our decisions on mistaken beliefs can lead to undesirable outcomes. Thus, one of the practical purposes of studying logic is to understand how to avoid moves or inferences that lead to error.

When we carefully examine the form of arguments, apart from the specific content of the arguments, we can more easily find problems with their structure which might otherwise be difficult to detect. We will see many examples of this strategy in the chapters which follow.

An inference is simply a move from one claim about the way things are (or the way things could be) to another.

By studying logic, we can learn how to improve our ability to draw legitimate inferences and to argue effectively. As we have seen, an argument involves a set of sentences that are offered in support of some claim or contention. As we study logic, we develop our ability to recognize that some inferences are good and some are bad.

3.3 Considering Alternatives

Once we have clarified what is at stake in a conversation or in a text and what its assumptions are, we will be in a position to begin evaluating the quality of the inferences in earnest. This task will become the principal focus of our study in later chapters. However, even in informal contexts, our analysis of an argument must pay attention to the quality of inferences.

At each point in the argument, where an assertion is made, we are entitled to consider counterexamples. For example, one simple practice to introduce into your daily practice of argumentation is to simply pause to ask a *what if. . .?* question. *What if*. . .questions are simply invitations to consider alternatives. Thus, another important feature of learning to reason well is the cultivation of our creative imagination. Excellence in reasoning requires the intelligent use of imagination. Imagination is what permits us to discover the right kinds of questions to ask in the course of an argument, what kinds of strategies to employ in a formal proof or in an informal argument. Imagination is vitally important insofar as it allows us to consider alternative ways things could be. If we were unable to consider alternative possibilities, our ability to think, to solve problems, and to argue effectively would be severely limited.

Not all of us are especially imaginative. If this is the case, how do we cultivate our imaginations? **The simplest way to stimulate our imagination and generate alternatives in the midst of an argument is to simply deny some assertion.** As an exercise, it is useful to deny some commitment that you may not have questioned.

> What if democracy is not the best form of government?
>
> What if marriage is harmful to society?
>
> What if happiness is not a worthy goal for human beings?

Simply considering the possibility that common beliefs of this kind are false is a useful way to get the imagination working. When it comes to specific arguments, our strategy might run along the following lines:

> Let's say that in the course of an argument concerning some ethical issue involving college athletics in the United States, one of the conversational partners mentions that college athletes should be treated just like any other student. We could open up a whole range of new considerations by considering the denial of that claim. What if it is not the case that student athletes deserve precisely the same treatment as nonathletic students? Once we begin to consider this possibility, a wide range of new options and lines of discussion can open. We might determine that the amateur status of college athletes is detrimental to their successes students that it is unfair not to financially remunerate student athletes for their work on behalf of the university, etc.

As we construct our own arguments, it is useful to keep a well-intentioned but highly critical opponent in mind as we write or speak. What this means, in effect, is that we ask ourselves whether there might be good reason to deny some step or assertion in our argument. Think of the cleverest person that you know, now imagine that he or she disagrees with what you say. What reasons might this person have for disagreeing with you?

Our friend, the imaginary opponent:

In the practice of argument analysis, it is an extremely useful practice to imagine that the argument has an opponent and that he or she is reading along with you as you study its structure. Imagine that this opponent is a clever person who denies the thesis of the argument. Imagine also that this clever opponent is, like you, a fair minded and reasonable person who is primarily interested in determining whether the argument provides good reason for him or her to change their view.

The very act of considering the possibility that I might be wrong helps me to see more clearly how I might strengthen my argument. In the extreme case, considerations of this kind might cause me to abandon my main contention. If I determine that the evidence and argument against my contention is stronger than my contention then I ought to simply abandon my thesis and adopt the contrary position.

Keeping your opponents in mind, as you develop your argument is a useful technique for strengthening your position or for coming to recognize that you ought to abandon your position. In either case the results are salutary.

In the same spirit, in discussions with others, serving as a constructive collaborator in the conversation rather than an opponent can help all involved pursue inquiry into more successful manner. In one's writing and conversations, it is useful to recognize that all of us are fallible; we are all susceptible to error. If we are genuinely interested in holding true beliefs and making decisions that are based on good reasons, then one's principal goal in a discussion is not to defeat our opponents in a verbal battle. Similarly, one's goal should not be to defend one's initial view and to defeat the views of one's opponent. Instead, the goal is to correct one's own errors and to make decisions for good reasons rather than bad ones.

Being charitable does not mean failing to be critical, instead it means engaging with the strongest possible version of the argument as presented by one's conversation partner. Likewise, recognizing that one is fallible does not mean deferring to the views of others. Instead, it means steering clear of a dogmatic attachment to one's own views and a recognition that all of us can improve our understanding.

Considering alternatives also allows us to stay focused on the actual premises and inferences in an argument rather than getting sidetracked in an unproductive way. For example, consider the following argument:

> Jacob:
>
> Athletes who are adults should not be penalized for using performance enhancing drugs. I think this because adults should have the right to put whatever substances they like into their bodies. We need to respect their autonomy.
>
> Jessica:
>
> You're wrong, organizations like the NBA and the NFL are private organizations and they can set whatever rules they like
>
> Lavalle:
>
> Performance enhancing drugs are potentially dangerous to your health. They shouldn't be allowed.
>
> Faisal:
>
> Jacob, you don't have the right to put anything into your body that might harm others. What about roid rage? The athletes might hurt other people.

Jessica and Lavalle are not addressing Jacob's reason for believing his position. Instead, they are offering reasons in support of the negation of his main contention. Faisal, by contrast, is taking issue with the reason Jacob believes what he does. Faisal's response might not convince Jacob to abandon his argument, he might continue to believe that the autonomy of adult athletes is more important than the potential harm of "roid rage" to others, or he might reject the empirical claim that performance enhancing drugs like anabolic steroids really increase the likelihood that an athlete will engage in violent behavior.

However, Faisal, unlike Jessica and Lavalle, has engaged Jacob's argument directly, simply via the method of considering alternatives. This shows that he has made the effort to understand the reasons that Jacob offers in support of his contention. By offering an alternative to the reason on offer, Faisal is actually helping Jacob to refine his reasoning and strengthen his argument. By contrast, Jessica and Lavalle are opening new lines of argument (against the main contention) in parallel with the original. Faisal's approach allows for a more focused and potentially productive joint investigation.

3.4 The Analysis of Dynamic Informal Arguments

Editorials in the newspaper, legal briefs, argumentative essays in disciplines like philosophy, and other written texts do not change in any significant ways over time. Written arguments are relatively stable records of an author's argument. However, the kinds of arguments that we encounter most frequently in daily life are quite different from the formal arguments we find in academic papers and in legal writing. Ordinary arguments generally include people whose views can change over the course of the discussion. This is an especially prominent feature of face-to-face conversations.

Conversations are dynamic and usually the reasoning of the participants in a discussion takes a very fluid form. Conversation partners can give different weight to different reasons over the course of time; discussants can jump from one sub-argument to another in the course of an argument, can be convinced to revise their assumptions, and can even modify their main contention as the argument progresses.

Ordinary arguments operate in a variety of messy ways and on multiple levels. Simply identifying the parts of arguments in ordinary conversation and decision making can sometimes be quite challenging. In these contexts, it takes considerable sensitivity to learn to recognize the parts of arguments and to evaluate the steps that arguments take.

The most important step in the analysis of ordinary arguments involves understanding the purpose, or main contention of an argument. This will not necessarily be obvious from the text of the argument alone. For example, the text might include no conclusion markers of the kind mentioned in Section 2.1. How are we to know the point of a conversation or argument? In addition to the importance of providing charitable interpretations, in order to find the main point of contention in an argument, it is also important to understand something about the context of the conversation. Consider the following example:

> Joining the conversation in progress:
>
> 5-year-old Child: "Can we just go and look at the dogs in the pound?"
> Parent: "I don't think we're ready to take care of a new dog right now."
> 5-year-old Child: "You're not ready to save a puppy from the pound?"
> Parent: "Not right now."
> 5-year-old Child: "So, when will you be ready to save a puppy?"
> Parent: "I don't know, maybe next spring."
> 5-year-old Child: "Oh" (with a pensive look)
> Parent: "OK, let's go" (grabbing car keys and heading to the local animal shelter)

The child in this case may not have intended to manipulate the parent or to present an argument. The child's questions may have been perfectly innocent. But we can interpret a fragment of a conversation like this as an argument. Even if the child was not deliberately presenting an argument to the parent, the conversation clearly contributed to the parent changing his mind about going to the dog pound. As such, the conversation had the effect of changing the parent's position with respect to the question of whether or not to go to the pound.

It might be the case that the parent changed his mind about the decision to go to pound irrationally, or for reasons that were not very good reasons. However, in order to evaluate the decision, it is helpful to think of this as an argument before asking whether the parent's decision was based on legitimate reasoning. In a case like this, our positive feelings about puppies probably overwhelm our ability to think clearly.

Of course, the problem with interpreting a conversation like this as an argument is that we have relatively little guidance as to what hidden premises are at work in the argument. Knowing more about the context and the intentions of the speakers would change our interpretation of the argument. If we knew, for example, that a beloved family pet had recently died, or that the parent was not in the habit of spoiling the child, it would change the set of hidden assumptions that we would find reasonable to assume.

While there is a way of reconstructing this fragment of the conversation as an argument, we often do not have enough evidence to decisively support one interpretation of the conversation over all other competing interpretations. However, independently of these considerations, it is important to reflect on our reasoning insofar as it influences our decisions

> **Insofar as arguments influence our decision making in important ways, we ought to try to evaluate the reasons given in support of the argument, and we must also be able to determine whether the moves that lead from those reasons to the main contention of the argument are warranted.**

For example, if you were the parent, imagine pausing before grabbing the car keys to consider whether you should decide to go to the pound. What kinds of questions would you ask yourself as you deliberated?

> Why was I originally disinclined to get a puppy?
>
> > Because I do not have the time or energy to care for a puppy.
>
> I changed my mind, but am I still concerned about the time, cost, and work involved in caring for a puppy?
>
> What are the reasons for me to change my position?
>
> Are they good reasons?

Consider a situation from your own life where a friend or family member convinced you to change your mind about some course of action. As you attempt to explain the reasoning that led you to change your mind, it is worth carefully considering your reasons for acting one way or another. As you think about this situation, consider a few questions about your reasoning:

> Why did you originally disagree?
>
> What were some of the assumptions that you shared with your friend or family member?
>
> What caused you to change your mind?
>
> > Was new evidence introduced to the discussion of which you were previously unaware?
> >
> > Were you unable to respond to specific counterarguments?
>
> Overall, were you right to be convinced?

Let's consider a conversation concerning some controversial topic. For example, if Jane claims that 16-year olds should be permitted to vote, we could ask her why she believes this claim. In response to the challenge, she will provide her reasons. We would then be in a better position to decide whether to agree with her or not. In a case like this, we already know what she is arguing for; what the main contention of her argument will be. Her argument is meant to lead us to accept her contention that 16-year olds should be permitted to vote.

It is likely that prior to listening to her reasoning, we will have our own opinion on the question. But if we are charitable and are genuinely interested in the truth, we will listen to her reasons and evaluate them as objectively as we can. Basically, we need to determine whether Jane's stated reasons for believing her claim convince us to agree with her?

If she is interested in convincing us to voluntarily assent to her contention, she may decide to present her reasons in the form of an argument. Ideally, she could respond with a set of supporting claims about evidence, some shared assumptions, and some inferences connecting her evidence and assumptions to her main claim. If her inferences are warranted, and if her assumptions and evidence are shared by her audience then her audience should accept her contention that 16-year olds should be permitted to vote.

We can think of an argument like Jane's as an extended answer to a *why*- question. When we ask why someone believes something we are asking them to provide their reasons for believing the claim and an account of why those reasons support their claim. The reasons and inferences she offers in support of her claim constitute the heart of her argument. In this case, an argument in support of her central contention might involve reference to the maturity or the cognitive capacities of teenagers, other rights and responsibilities that pertain to 16-year olds, examples of other situations in which 16-year olds can vote, etc.

Of course, in ordinary conversation, arguments are usually highly informal, with many hidden, or assumed steps and assumptions. It is not always clear what the inferences are, what connection they have to the main claim of the argument, and sometimes it might not even be clear what the main claim of the argument is. In our case, we know that Jane is arguing for a specific conclusion, so this certainly helps us to evaluate her argument. However, let's spend a moment listening in on her conversation to see how we can begin to identify her argument.

We overhear a conversation already in progress where Jane attempts to justify her claim in response to objections from Sue and Carlos.

Jane: ". . .But actually, I think 16 year olds *should* be permitted to vote" [This will be her central contention; the main claim of her argument]

Sue: "You can't be serious, why do you think that?" [Her interlocutor is asking her to provide an argument for her claim]

Jane: "Well, teenagers are already trusted with many adult responsibilities and if they have a job, they have to pay taxes, so it's only fair that they have a voice in how those taxes are spent." [At this point, Jane has presented some evidence for her view, has introduced a number of unstated assumptions and has made one explicit inference]

Sue: "But surely most 16 year olds don't know enough about politics to be informed voters" [Sue has raised an objection to her central claim]
Jane: "Look, there are plenty of people over 18 who don't know enough about politics and they have the right to vote" [Jane responds to that objection]
Carlos: "Maybe you're right, I know plenty of foolish 40 year olds, but from my point of view, I think high school students should be protected from the additional responsibility of voting so that they can concentrate on their education and on enjoying their teenage years" [Carlos concedes Jane's point—we should not disenfranchise ignorant people— but raises another reason not to accept the main claim]
Jane: "I don't agree. After all, 17 year olds in the United States can serve in the military, it's not like they're protected from responsibilities to their country. As for school work, maybe they would take their studies more seriously if they were paying attention to important political questions" [Jane responds to that objection]
Sue: "I don't think 16 year olds are mature enough to be trusted with voting" [Sue raises another objection to her central claim]
Jane: "In the United States, teenagers are trusted with driver's licenses and in most states, the legal age of consent for sex is between 16 and 18. So if we think that they're old enough to drive and have sex, then they're mature enough to vote." [Jane responds to that objection]
Carlos: "Jane, it's just unrealistic to expect that 16 year olds will be allowed to vote in this country, we shouldn't even be arguing about this" [Carlos questions the whole point of the discussion]
Jane: "Even if you're right Carlos, the fact that people over 18 won't allow 16 year olds to vote, doesn't mean that they *shouldn't* be allowed to vote." [Jane defends the debate and attempts to refocus her conversation partners on her main contention]

In ordinary conversation, a great deal often goes unsaid and it is sometimes difficult to follow precisely what the steps of the argument really are. We might even wonder whether it is actually correct to interpret some conversations as arguments. Even if it seems appropriate to interpret a conversation as an argument, conversations often take directions that have no bearing on the main topic of the argument, derailing the argument, or distracting the participants. It is often useful to separate out the relevant portions of a written or spoken argument in order to discover what the main contention of the argument might be, to understand the evidence or reasons offered in support of the main contention and to determine the quality of the inferences involved in the argument.

An important feature of Jane's argument, in the imagined discussion above, is her **response to objections**. Carlos and Sue don't accept Jane's contention that 16-year olds should be granted voting rights and in our imagined case, Jane spends most of her time responding to their objections. This is understandable. After all, as we have seen, for the sake of efficiency, certain sub-arguments for her thesis can be safely neglected and certain pieces of background knowledge or points of agreement can be assumed given her familiarity with the views and background of her audience. She concentrates instead on the points of disagreement.

There are many aspects of ordinary reasoning and argumentation that are sensitive to context and they can seem very far from the ideals of formal logic. For example, it is generally the case that discussions like this never reach final resolution. The participants may simply get sidetracked, lose interest in the topic, or simply run out of time. Real conversations and debates are generally messy and inconclusive. In our less charitable moments, we might conclude that real participants in arguments often demonstrate less of an interest in a fair evaluation of the truth of the matter at hand than we might prefer.

We hope that when presented with evidence on both sides of some controversial issue, we will be inclined to objectively evaluate the evidence. We hope that we are open to changing our view on the matter if such a change is warranted by the evidence. Similarly, when presented with good arguments that demonstrate the error of our position we hope that, we would have the courage and intellectual maturity to change our position.

Unfortunately, there is strong empirical evidence showing that when exposed to evidence on both sides of some controversial dispute, subjects in psychological experiments tend to become **more convinced** of the correctness of their own view, even when instructed to be even-handed in the evaluation of evidence. This phenomenon is known as **attitude polarization** or **group polarization**. Psychologists have found that conversations concerning controversial or contentious topics frequently generate more extreme disagreements rather than more understanding or conciliation. This is one of the negative consequences of **confirmation bias**: the tendency to ignore evidence that does not support one's own beliefs and to put excessive weight on evidence that does support one's own beliefs. Our existing commitments and strongly held views distort our ability to evenhandedly assess new evidence or information. Thus, confirmation bias tends to make us illegitimately reinforce our confidence in our cherished beliefs while downplaying evidence that should cause us to change our minds. In later chapters, we will discuss these so-called **cognitive biases** in greater detail. As we shall see, psychologists have argued that bias makes it difficult for us to consider arguments in a fair and reasonable manner.[1]

In this context, it is important to recognize that dynamic and informal arguments are often shaped by biases. We should be alert especially to alternatives that go unconsidered and to possible objections to which no responses are given.

3.5 Being Fair-Minded versus Wanting to Win

In situations where it is genuinely important to determine the truth or falsity of some proposition, it is useful to approach conversations and arguments in an informal context with a spirit of openness and generosity. Often, we can be in a position to assist another person in determining precisely what his or her commitments are. Our partners in conversation may not actually have a fixed position fully articulated or clearly in mind when they engage in an argument with us. As we help our conversational partner determine his or her position in an argument, we are contributing to sharpening and clarifying our own

[1] One classic source for evidence concerning attitude polarization is Lord, Charles G., Lee Ross, and Mark R. Lepper. "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence." *Journal of Personality and Social Psychology* 37, 11 (1979): 2098.

position. In most informal contexts, a directly adversarial approach wherein we seek to defeat our opponents does not serve the goal of pursuing the truth. Obviously, there are cases where it will be necessary to show our conversational partners the error of their ways. However, we ought to be cautious before presuming that we are right.

Many of the things we know or understand very well are not easily put into words. Think, for example, of the ability to ride a bicycle, expertise in sports, sensitivity to social situations, and the like. It would be a mistake not to take the views of experts in matters like this into account simply because they were unable to articulate a clear thesis from the outset. If a great chef or a juggler gives you advice on their craft, even though they are unable to justify their reasons for giving the advice, it is prudent to take their advice seriously. Thus, it may be the case that someone understands a problem or situation very well without being able to clearly articulate a view that they could defend in a rigorous argument.

Ideally, when philosophers are involved in arguments, we are less interested in defeating our adversary and more interested in the pursuit of truth. In this sense, most philosophers see a productive argument is a cooperative endeavor. Arguments, at least as philosophers view them, should be a clash of ideas and an examination of reasons rather than a conflict between persons. On this view, the purpose of argumentation is to arrive at a clearer picture of the truth, not to demonstrate one's superiority over an opponent. Notice that these are normative claims rather than descriptive ones.

The origins of our ability to reason and the uses to which we put this ability are probably not very noble. Specifically, there is some evidence to suggest that our ability to reason did not evolve in order to assist us in the pursuit of truth. Hugo Mercier and Dan Sperber have recently defended the view that reason probably evolved for the purposes of convincing other people to adopt courses of action preferred by the arguer.[2] The ability to convince others would certainly have been a very useful ability to cultivate over the course of our species' evolutionary history.

[2] Mercier, Hugo and Dan Sperber. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory (June 26, 2010)." *Behavioral and Brain Sciences* 34(2), 57–74. Available at SSRN: http://ssrn.com/abstract=1698090.

However, as Mercier and Sperber would undoubtedly agree, we should not let the humble origins of our ability to reason lead us to neglect its central role in the pursuit of truth. While reasoning might not have evolved for the purposes of pursuing the truth, it is the best tool we currently have for doing so.

Cooperation and a spirit of open-minded inquiry are essential to the kinds of arguments that concern philosophers and others interested in distinguishing true beliefs from false ones. While one can consider the virtues of an argument in isolation from others, arguments generally involve a community of people. Arguments between people of good will are generally a good way of improving our understanding of some topic. Defending and justifying our views in the company of sophisticated adversaries force us to examine the reasons for our beliefs. Discussion and argument allow us the opportunity to evaluate and revise our own views and make it easier for us to see when we are mistaken. If our goal is the pursuit of truth, critical debate is a good method to adopt. Arguments may have an adversarial flavor; they may sometimes feel like battles, but genuine arguments would be impossible without a great deal of common ground and cooperation.

When two or more persons engage in the kinds of constructive arguments that we have in mind here, they must

- share enough common ground to know that they disagree.
- share some common presuppositions and a common vocabulary.
- hope to convince their opponent without coercing them.

Notice that these three points are fully consonant with an evolutionary explanation that emphasizes the origins of reasoning in our natural human tendency to attempt to manipulate others.

If the pursuit of truth is one's principal goal, then in the face of strong evidence,

Arguments may have an adversarial flavor; they may sometimes feel like battles, but genuine arguments would be impossible without a great deal of common ground and cooperation.

or
when confronted with good reasons to abandon one's claims, one should be willing to drop the claims one had previously accepted. Our tendency as human beings, as we shall see in more detail in later chapters, is to remain loyal to our preferred views for as long as possible. However, we should beware of the difference between being loyal to some set of views and clinging to them stubbornly or even unreasonably in light of good evidence against one's views.

One should maintain an attitude of humility, and recognize one's own limits as a thinker. The virtues of humility and fairness are closely related to the principal of charity with respect to the claims and arguments of others. Unless we have good evidence that leads us to think otherwise, we should take the position of our adversary as seriously as we reasonably can. Unless we have excellent evidence to the contrary, we should assume that our opponent is a rational and well-intentioned person who believes that she has good reasons for holding the positions she does.

It is usually a bad argumentative strategy to dismiss one's opponent as a fool or as an immoral person. **Even if it turns out that your opponent actually is a bad or foolish person,** providing a charitable interpretation allows us the opportunity to test our own views against the strongest possible adversary.

3.6 Persuasiveness: The Art and Science of Rhetoric

The real strength of an argument depends on the truth of its underlying assumptions and on the quality of the inferences drawn from those assumptions. The other important dimension of arguments is their *apparent* strength. Some faulty arguments *appear* to be strong and can persuade an unwitting audience for bad reasons.

As mentioned above, the principal focus of this book is the form of arguments. However, an argument can be formally impeccable and yet fail to convince or persuade an audience of its conclusion. If one's goal is to be persuasive, the formal correctness of one's argument is not sufficient. This is why, in informal settings, we need to pay at least some attention to persuasiveness.

Whether an argument is convincing or not depends to a great extent on the nature of the audience that is being addressed. The study of how to persuade an audience is known as rhetoric and in most universities rhetoric is usually studied and taught by nonphilosophers. By contrast, logic is usually taught in Philosophy or Mathematics departments.

We philosophers sometimes dismiss rhetoric as focusing on superficial matters of presentation and style rather than on the genuine pursuit of truth. Rhetoric is sometimes criticized by philosophers because they believe that it ignores the truth of claims or the validity of arguments. More seriously, philosophers are sometimes suspicious of the ethical status of rhetoric insofar as it is associated with forcing or tricking an audience into accepting a claim in a morally objectionable way. Detractors will point to the misuse of rhetorical techniques in manipulative journalism, advertising campaigns, or in political propaganda.

> Concern with the dangers of rhetoric has been a prominent feature of philosophy. This is due in large part to Plato's criticism of the sophists. In ancient Greece, the sophists were paid teachers of the art of rhetoric and persuasion. Because Athenians argued their own cases in court (there were no lawyers to represent them), it was important for wealthy Athenians to learn how to persuade their audiences. In many of his dialogues, Plato accused the sophists of being unconcerned with the pursuit of truth. Through their art, Plato charged, they were able to "make the weaker argument appear the stronger."

In spite of the long history of antagonism between philosophy and rhetoric, there is a lively study of informal argumentation whose practitioners are not trying to be coercive or unethical. Instead, the goal of the academic study of informal argumentation, let's call it *the ethically sensitive study of rhetoric* is to understand how one gains voluntary assent to a claim or conclusion rather than how one forces the audience into agreement. Learning how context, interests, style, and presentation influence the reception of an argument does not necessarily lead to the misuse of that knowledge for coercive purposes.

Aristotle understood rhetoric to be

> "the faculty of observing in any given case the available means of persuasion."

The Aristotelian view is that even if you have a valid argument for a true claim, these factors may not be enough to convince an audience. They may not care enough about the topic to pay attention to what is being argued, or the manner in which the argument is presented may be offensive or annoying to the audience. If we care about effective argumentation, then, as Aristotle said, it is necessary to understand how to engage with the interests and capacities of our audience in order to persuade them to accept our claims.

Imagine that I am trying to convince an audience that governments should oblige citizens to perform 1 year of national service in the military. If I hope to argue effectively, then the way I argue for my contention will depend, at least to some extent, on the nature of my audience. Distinct groups—for example, teenagers and elderly people, political conservatives and political moderates, and anarchists and communists—are likely to be more amenable to different kinds of evidence or different lines of argument. Effective and persuasive reasoning involves taking these differences into account and modifying our arguments accordingly.

In this book, we will return, occasionally to the question of the persuasiveness of arguments. However, following the philosophical tradition established by Plato, our principal focus in this book will be on validity and soundness and on developing the kind of intellectual virtues which support the pursuit of truth.

3.7 Guidelines for Argument Analysis in Informal Contexts

Argument analysis can be conducted at varying levels of depth. However, we already have the beginnings of a simple method for approaching an argument.

Argument analysis begins with the following steps:
First Step:
Charitably sketch the main skeleton of the argument.

A. Find the main contention (what is the point of the argument?)
B. Find the principal reasons that are offered in support of the main contention. (Why does the author think we should agree with the main contention?)
C. Find the premises of the argument. (What assumptions serve as the basis for the author's argument?)
D. Find points where the author addresses objections to the argument. (Does the author successfully reply to opponents of the argument?)

Second Step:
Imagine a reasonable and well-informed person reading the argument who does not agree with the main contention.

A. How might she respond to reasons offered in support of the main contention?
B. Would she accept the premises of the argument?
    a. Is there sufficient evidence to accept the premises?
    b. Are the hidden premises of the argument acceptable?
C. Would she raise objections which are not addressed by the author?
    a. Does the argument presented by the author contain an obvious way to handle those objections?

Third Step:
Detailed consideration of the inferences:
A. Are the inferences fallacious?
    a. Do they contain formal errors in logic or probabilistic reasoning?
    b. Do the inferences illegitimately undermine inquiry?
    c. Do they reveal the negative influence of cognitive biases of one kind or another?

Exercise for Chapter 3

Reading Newspaper Editorials and Opinion Pieces

1. Newspaper editorials provide useful material for practicing argument analysis. They are usually weak arguments, written in a hurry, and often very poorly reasoned. In this exercise, you should begin by reading an editorial carefully before beginning to extract the structure of the argument.

(a) Begin by determining the main contention or conclusion of the editorial. Are there alternative plausible interpretations of the text that would have different main contentions? If so, why did you settle on the one you did?

(b) What parts of the text are genuinely relevant to the argument? For the instances, you identify as irrelevant, explain how you determined that piece of text to be irrelevant to the argument.

(c) What are the reasons that the author gives in support of their conclusion? Can you identify inferences or moves in the argument?

(d) Does the argument depend on unstated assumptions? What are they? Is the author right to believe that their audience will share their assumptions?

(e) Does the author consider objections to their reasoning? What objections does the author fail to consider?

4 From Common Sense to Formal Reasoning

4.1 Logic as a Normative Science

Psychology and neuroscience provide our best hope for useful descriptions and explanations of how the human brain works. The goal of these sciences is to help us to understand our brains and behavior. Experience and experimentation has shown us that damage to the brain can systematically change our cognitive, emotional, and behavioral capacities. For over 200 years, we have had strong evidence that the brain is involved in the life of the mind. We know also that our brain chemistry can be influenced in ways that change our moods, our behavior, and our thinking. Today, we know a great deal about the cellular and subcellular details of the nervous system and thanks to experimental psychology we understand a great deal about patterns in human behavior. The messy reality of embodied human life is the focus of attention for valuable scientific research.

It is likely to surprise most readers to learn that logicians have had almost no interest in the way human brains work. Research-level logicians are, for the most part, uninterested in how human beings happen to think through problems or make decisions. There are good reasons for their lack of interest.

Consider, for example, that embodied human cognition is limited in very distinctive ways due to its biological basis. Our short-term memory is not very large, our cells need oxygen to survive, many of us need large quantities of coffee to get started in the morning, our attention wanders, etc. Humans think using brains that have limits and idiosyncrasies that logicians simply do not care about. These facts are directly relevant to having embodied minds like ours, but coffee, oxygen, and attention are not relevant to logic itself. This is because logic aims to help us to understand some of the central characteristics of reasoning per se rather than reasoning as we happen to engage in it. Logicians are interested in the nature of excellent reasoning, and not in the nature of particular reasoners; they do not concern themselves with whether a thinker is meat-based, silicon-based, or takes some other form. Logic shows us how we *should* reason if we wish to achieve formal correctness and not how our brains or minds happen to reason most of the time.

Because it is focused on excellent rather than actual thinking, logicians sometimes say that their discipline is **normative** rather than descriptive. A **norm** is a rule or a principle that serves as a guide for action or judgment. Logicians consider the norms that characterize good reasoning. When we violate these norms, reason goes wrong.

> Logic is normative rather than descriptive. This means that it is the study of how we *ought* to reason rather than a description of how we actually reason. A norm is a rule or a principle which serves as a guide for action or judgment. Logic provides the rules or principles for distinguishing good reasoning from bad.

Like ethics, logic can play a normative role in our decision making. Logic can be applied to arguments and inferences in roughly the same way that ethical standards and rules can be applied to actions.

One way to think about norms or standards is to reflect on some simple cases that are obvious violations. From there, we can try to determine whether there are certain characteristics of those cases that can be taken as models when we are reasoning about other, more complicated cases. As we saw in Chapter 1, we derive our basic logical norms and capacities from a common sense ability to recognize when an argument has gone wrong. Consider how easily we can recognize obvious contradictions in an argument, clear failures in relevance, obvious nonsequiturs and the like:

> It's true that I haven't paid the taxes I owe for 10 years, but I swear, I'm a law-abiding citizen
>
> Because Socrates was ugly, he was a good philosopher
>
> I wouldn't trust her to watch my purse, she's Belgian.
>
> If the state permits same sex couples to marry, why not allow people to marry household appliances?
>
> People who oppose gun control just want kids to die.

We might not be able to say exactly what is wrong with these pieces of bad reasoning, but most of us can tell that there is a problem in each one. As we saw in Chapter 1 in Section 1.4, we don't need a course in logic to see the problems with statements of this kind. **The trouble is, when it comes to the analysis of arguments, common sense is reliable <u>only in the most obvious and basic cases</u>.**

More disturbingly, common sense reasoning tends to be lazy and as we shall see, we habitually take shortcuts in thinking which can systematically mislead us. In examples like the ones just considered, we rely on common sense to tell us that there is a problem with this argument, but the cases we encounter will not always be so easy. While we will begin from our common sense understanding of errors in reasoning, we must supplement it with formal methods.

4.2 Our First Fallacy: Affirming the Consequent

In a moment, we will consider a very simple example of reasoning badly. Fallacious arguments can take a wide range of forms. Over time the most common kinds of fallacy have gotten their own names. Our first example is an instance of a pattern of bad reasoning that we call ***the fallacy of affirming the consequent***. The fallacy of affirming the consequent is an example of a formal error in reasoning; its victims mistake a bad pattern of reasoning for a good one. Some instances of this fallacy are obviously problematic and would never trick us. Let's look at an example of the fallacy that is easy to recognize as bad:

> Premise 1: If Einstein invented the driverless car, then he is smart.
> Premise 2: Einstein is smart,
> Conclusion: Einstein invented the driverless car.

Here, you are likely to suspect that reasoning has gone wrong because you know that Einstein did not invent the driverless car. Your awareness of the factual error in the conclusion probably causes you to doubt the reasoning that produced it. However, the factual error is completely irrelevant to the logical error in the reasoning here. In fact, we are just lucky to have an obvious factual error in the conclusion. The real problem with the reasoning here is not the fact that the conclusion is false but with the pattern of reasoning that led to the conclusion. As we will see, other cases are less obvious. Here's another obvious case that has the same problematic structure as the Einstein example above:

> If the Queen owns the crown jewels, she does not have to worry about funding her retirement. It is true that she does not have to worry about her retirement, so she must own the crown jewels.

The Queen case is slightly less obvious than the Einstein case, but you can see that there is a problem here too. It is worth moving slowly through a formally equivalent example to these cases in order to understand as precisely as possible what the problem is with the form of a pattern of reasoning like this one.

Let's follow the reasoning that leads Sam to give a piece of advice to his friend Karla. As we will see, he **illegitimately moves from a pair of true sentences to his conclusion by a formally incorrect method**. His conclusion supports his decision to give Karla an unwanted and unnecessary piece of advice.

Here is the scenario:

> Sam knows that whenever Karla stays up late at night she's tired the next day.
>
> He sees that Karla is tired this morning.
>
> He concludes that she stayed up late the previous night and advises her that she ought to go to bed earlier in future.

Let's assume that Sam's reasoning is based on two secure pieces of knowledge;

> knowledge of the fact that Karla is tired this morning
>
> and
>
> knowledge of the rule that if Karla stays up late at night, she will be tired the next day.

From these two sentences, Sam concludes (incorrectly) that she must have stayed up late the previous night. This conclusion informs his decision to advise her to go to bed earlier in future. The problem with his reasoning becomes clear when she tells him the following:

> "I didn't stay up late last night. I got up very early this morning to finish my logic homework."

From the fact that whenever Karla stays up late at night she's tired the next day and the fact that Karla is tired today, Sam **should not have concluded that she *must*** have stayed up late the previous night. He mistakenly ignored the possibility that her tiredness could have been caused by factors other than staying up too late. He has used an illegitimate pattern of reasoning.

The reason that it is a faulty piece of reasoning is that Sam does not notice that there are **counterexamples** that undermine his argument. Given what he knows, as stated in the premises, he is not entitled to ignore other possible causes for Karla's tiredness. For example, she might have run a marathon or she might have a disease, or as she reports, she might have gotten up early to finish her logic homework. Just because she is tired, it's not necessarily true that she stayed up late the night before.

Premise:	When Karla stays up late at night she's tired the next day.
Premise:	Karla is tired this morning.
Conclusion:	Karla stayed up late the previous night.

Some counterexamples to the conclusion:
Karla did not stay up late the previous night, instead she woke up very early.
Karla did not stay up late the previous night, instead she ran a marathon.
Karla did not stay up late the previous night, instead she is ill.

Counterexamples are sentences that could be true given what we know from the premises.
If these counterexamples could be true, then given the premises, Sam's conclusion is not necessarily true.

Sam's conclusion was based on the strong **association** that he had in mind between Karla's tiredness and the fact that when she stays up late, she is tired the next day. This strong association overshadowed other possibilities and led him astray. Fallacies are bad arguments precisely because they lead us astray and cause us to reason badly. Sam failed to consider the possibility that his conclusion could be false while the premises of his argument are true. The power of the strong association between being tired and staying up late led him to conclude that she must have stayed up late. Notice that in doing so, he is illegitimately excluding the counterexamples that we considered above.

Let's be charitable readers of his inference for a moment. On its face, the inference itself is faulty, but could there be a way in which Sam could be justified in saying what he said?

Imagine a scenario in which he is asked simply whether she stayed up late. Imagine also that he knows that when she stays up late then she is tired and he knows that she is tired. Now imagine also that he knows an additional piece of information, namely that

> *The only way she could be tired is the circumstance in which she stays up late.*

If he had this additional piece of information, then he would have been justified in making his inference. Notice that the additional piece of information would close off other ways that she might have become tired, thereby ensuring that he could not be wrong.

Notice that given what he knew, Sam would have been committing a fallacy **even if it had been true** that Karla had stayed up late the previous night. In this case, we noticed the problem with his reasoning because he was led to conclude something false, namely that she had stayed up late the previous night. But as we shall see, fallacies like this one are bad even if the sentences that figure in the fallacious pattern of reasoning are true. **Patterns of reasoning can be evaluated independently of the actual truth and falsity of the sentences involved**.

> In Sam's defense, he might argue that the fact that Karla's tiredness provides him some reason to believe that she stayed up late the night before. Clearly he was wrong to say that she *must have* stayed up late, but perhaps he is entitled to say, more modestly that her tiredness is evidence of some kind. Later in the book, we will consider some reasoning about probabilities and will revisit the question of whether Sam might have been right to think that Karla's tiredness supports the judgment that she *probably* stayed up late.

Let's imagine that in fact, she *did* stay up late the previous night. Even so, she is entitled to respond:

> "Hey, you don't know for sure that I stayed up late last night, for all you know I could have run a marathon yesterday."

Independently of the truth or falsity of the conclusion, her criticism of his reasoning is still legitimate. Whether she is entitled to object to his reasoning is independent of the question of what time people should go to sleep and when she happened to go to sleep. The problem with Sam's reasoning is the result of **its form** and not the truth or falsity of the sentences themselves.

Many of the good patterns of reasoning that we frequently encounter also have names. As we shall see, following good patterns of inferences makes one less susceptible to formal errors of the kind that Sam committed. The following good piece of reasoning is an example of a pattern that is traditionally known as ***modus ponens***:

> Whenever Karla stays up late at night she's tired the next day.
> Karla stays up late at night.
> Therefore, Karla is tired the next day.

If it is true that whenever Karla stays up late at night, she is tired the next day and if she stays up late at night, **you can conclude with absolute certainty** that Karla is tired the next day. If one accepts the premises and uses this legitimate pattern of inference, it turns out that **there are no counterexamples to the conclusion**.

Notice that by accepting the first claim (whenever Karla stays up late at night she's tired the next day) as true without any exceptions, one has made a very strong claim indeed. The first premise denies the possibility that she might stay up late and not be tired the next day because of some medical treatment or by drinking lots of coffee. Nevertheless, while the premises might be dubitable, *modus ponens* itself is an example of a rock solid pattern of reasoning, and in later chapters we will demonstrate why good patterns can be relied upon with absolute certainty.

It is somewhat easier to see the form that this legitimate pattern of reasoning takes once we reduce it to its bare bones. This is what it looks like when we begin to strip *modus ponens* down to its formal structure. We begin by letting letters stand for sentences. In our case, we can let italicized lower case letters "*a*" and "*b*" serve as **variables** that stand in for sentences in the following way

"*A*" replaces "Karla stays up late at night"

and

"*B*" replaces "Karla is tired the next day"

What is a variable?

In basic mathematics letters (*x*, *y*, *z*, etc.) often serve as variables.

A variable represents some unknown or unspecified numerical value in a problem or an equation. For example, one might be given an equation like

$3x + 1 = 10$

and asked to solve for x. A little basic algebra lets us see that in this case $x = 3$.

In more advanced mathematics, variables can stand for other kinds of objects, not just numbers, but vectors, functions, and matrices.

At this stage in our study of logic, **variables will represent declarative sentences**.

Later, we will use variables to represent names, properties, and relations.

At this point, we are not talking about Karla, sleep, and staying up late. Once we have replaced the sentences with sentence variables, we can see the form of *modus ponens* more clearly:

Modus ponens

If *A* then *B*

A

therefore *B*

The particular content of the sentences is not important to the form of the argument. The sentence variables in this example allow us to ignore the meaning of the particular sentences, focusing instead on the function of logical phrases like "if . . . then," "therefore," "and," "or," and others. These logical phrases and terms are the bones and joints that hold the skeleton of logical form together. Instead of focusing on the flesh of the arguments, the true and false sentences, with all their strong associations and emotional resonance, we use variables to strip the argument down to its bare skeleton.

As we begin formalizing our arguments in this simple way, we can already compare this valid pattern of reasoning to the fallacy of affirming the consequent which takes the following form:

Fallacy of Affirming the Consequent

If *A* then *B*
B
therefore *A*

As we continue our study, we will examine the reasons for calling argument patterns legitimate or illegitimate. For now, we just need to recognize that as a matter of common sense, some ways of arguing are good and some not so good. In the case of *modus ponens*, we can see that the conclusion *follows from* the premises. Whereas when we see examples of the fallacy of affirming the consequent stripped down to their formal skeletons it is clear to common sense that the conclusion does not follow from the premises.

4.3 Deductive Reasoning

Logicians are concerned with how sentences *follow from* or are *implied by* some other sentences. Sometimes we talk about a sentence or a thought being a *logical consequence* of some other set of sentences or thoughts. If a sentence is a logical consequence of a previous sentence, then we can be confident that moving from the first sentence to the second is warranted.

Logically correct patterns of reasoning are ***truth preserving***. By calling "truth preserving," we mean that if we start out with some set of true sentences, the rules or patterns of good reasoning will never lead us to affirm a false sentence; reasoning in a logically correct manner will not lead us astray.

Some instances of logically correct patterns of reasoning:

Some clouds are not composed of water vapor; therefore, not all clouds are composed of water vapor.

Jen and Jan are mammals; therefore, Jen is a mammal.

Roger is chordate or Data is fleshy and Data is not fleshy therefore Roger is chordate.

All dogs are canines; therefore, some dogs are canines.

Danusha is in Texas; therefore, Danusha is in Texas or Jupiter is inhabited by Lopers.

Some of these will seem obvious to you already, some perhaps less so. In these examples, the claim following after the "therefore" follows logically from the claim preceding the "therefore." The idea of some sentences *following from* or *being implied by* others should puzzle you a bit. On the one hand, there is something intuitively obvious about the special relationship between the conclusion and the premises in *modus ponens*, for example, but understanding the nature of logical consequence involves deep philosophical challenges and will require more than ordinary common sense.

The idea that certain sentences or thoughts follow from or are implied by other sentences or thoughts leads logicians to pay special attention to **deductive reasoning**. Specifically, one of the tasks of logic has involved the goal of systematically understanding how to determine whether a sentence really does follow logically from some other sentences. Deductive reasoning involves sequences of thoughts or sentences that are connected by inferences. Take the following simple piece of deductive reasoning:

If Harry is in El Paso, then Harry will have a lizard-skin wallet.	⇦ PREMISE
Harry does not have a lizard-skin wallet.	⇦ PREMISE
Therefore, Harry is not in El Paso	⇦ CONCLUSION

As we shall see, this is a formally correct or valid piece of reasoning which moves from two assumptions to a conclusion. In this case, we would say that the conclusion follows from the premises or is a logical consequence of the premises.

> If it were true that
>
> If Harry is in El Paso, then Harry will have a lizard-skin wallet.
>
> and if it were true that
>
> Harry does not have a lizard-skin wallet.
>
> then you would know (without having to search the city) that the conclusion
>
> Harry is not in El Paso.
>
> is true.

Let's consider another example of a piece of deductive reasoning in a more informal context:

I know that Xotchil really likes Joy Division because she either likes to listen to them or she's pretending to like them so that she can impress her pretentious friends and I know that she is not trying to impress her pretentious friends.

Notice that in this little argument, the conclusion **I know that Xotchil likes Joy Division** is stated prior to providing the argument. The conclusion of an argument is not always the final sentence in a string of sentences. Here, the speaker presents her reasons for believing that Xotchil likes Joy Division.

The premises of this little argument are the two sentences

Xotchil either likes Joy Division or she's trying to impress her pretentious friends

and

Xotchil is not trying to impress her pretentious friends

And the conclusion is:

Xotchil likes Joy Division

The conclusion of the argument is that Xotchil likes Joy Division. Intuitively, you can probably already see that whenever one accepts the truth of the premises in this argument, one is compelled (somehow) to accept the truth of its conclusion. One might question whether our speaker should accept the truth of these assumptions, but given that the premises are true, it is incorrect to reject our speaker's conclusion. But how did this reasoning move from the premises to the conclusion and why do we feel so confident that these moves are legitimate?

In this case, her inference followed a legitimate pattern known as *deductive syllogism* (sometimes abbreviated as DS). This is one of many legitimate patterns of inference that we will study more in later chapters. Using common sense you can already see how it works. If you know that either *A or B* is true and you know that *B* is false, you can conclude that *A* is true. Alternatively, if you know that *A or B* is true and you know that *A* is false, you can conclude that *B* is true.

As we shall see, this piece of reasoning follows a rule of inference known as *disjunctive syllogism*. We'll see more of disjunctive syllogism and rules like it in later chapters. Peeking ahead a bit, we will see how patterns or rules like *disjunctive syllogism* can be symbolized. The italicized letters "*P*" and "*Q*" can stand for any sentences, the "¬" is the symbol for negation or "not," and the "∨" will stand for "or."

DISJUNCTIVE SYLLOGISM (DS)

$p \vee q$

$\neg p$

therefore

q

Given any premises or set of assumptions, some conclusions *must* follow; they follow necessarily. **If some conclusion follows logically from the premises, then in all possible circumstances where the first and second premises are true, the conclusion is true**.

> Instead of talking about possible circumstances or scenarios, many philosophers would say that in any *possible world* where the premises of an argument are true, the conclusion of a logically correct piece of deductive reasoning based on those premises is also true.
>
> A possible world is a total way things could be. The actual universe is one way that things could be; it is one possible world, but believers in possible worlds think that there are other ways for universes to be. The actual world is one, among many, total possible way things could be.

Good reasoning involves deduction. In the cases we have seen so far, you have probably found it easy to follow the steps in the deductions presented. However, there are a range of contexts in which human reasoners fail systematically to grasp deductive relationships. In order to see how we fail, let's explore how we solve problems using deduction. For example, simple logic puzzles of the following form involve your ability to understand deductive relationships:

> Imagine that the population of three Midwestern cities; Des Moines, Wichita, and Columbus is relevant to your business. You don't have access to the Internet, but you know that their populations are 822,553, 207,510, and 386,552. You know that Des Moines has fewer than 300,000 people and that Wichita is not the largest. So, what is the population of Wichita?

You can quickly determine the answer to a puzzle like this by deduction. Puzzles of this kind involve keeping track of the space of possible options, what you know from the instructions, and figuring out what is and is not ruled out by that information. The space of possibilities in this case is simply the list of cities and the three possible numbers.

Since you know that Des Moines must have fewer than 300,000 the only option for it is 207,510. You can represent the space of possibilities for this problem as follows:

	207,510	386,552	822,553
Wichita	No		
Des Moines	Yes	No	No
Columbus	No		

Given you know that Wichita is not the largest,

	207,510	386,552	822,553
Wichita	No		No
Des Moines	Yes	No	No
Columbus	No		

then it either has a population of 207,510 or 386,552, but since Des Moines has a population of 207,510, then it must have a population of 386,552

	207,510	386,552	822,553
Wichita	No	Yes	No
Des Moines	Yes	No	No
Columbus	No	No	

And of course, Columbus' population must be 822,553 since that's the only remaining place in the space of possible populations.

	207,510	386,552	822,553
Wichita	No	Yes	No
Des Moines	Yes	No	No
Columbus	No	No	Yes

Most of us are pretty good at finding solutions to problems of this kind. We can determine how the various pieces of information fit in the space of possible options and we can tell what options they exclude.

Logic puzzles like these feature in assessments of critical thinking ability in various standardized tests administered for entry into graduate and professional schools in the United States. Such questions can become much more complicated than the example given above, but their basic structure is similar. They outline a space of possible relationships to be filled in, some information is given, and the challenge is to determine what to conclude about the relationships given the information provided.

With a little practice, the kind of logic puzzles that appear on standardized tests can be mastered by most students. However, other simple problems related to deduction are far more challenging.

4.4 Wason's Task and the Limits of Our Common Sense Deductive Abilities

In 1966, the psychologist Peter Wason devised a simple experiment to test the ability of subjects to reason using simple rules.[1] Specifically, Wason's task is a logic puzzle that tests the ability to reason about an *if . . . then* relationship. The test takes roughly the following form: Participants are presented with four cards. They are told to assume that each card has a number on one side and a color on the other. They are then given a rule:

Rule: *if a card has a 7 on one side, then the other side is blue*.

Four cards are placed on the table in front of the participants and they are asked the following:

In order to judge whether this rule is true, which cards would you need to turn over?

[1] Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20(3):273–281.

Green	Blue	7	5
Card #1	Card #2	Card #3	Card #4

Testing the rule proves trickier than most participants realize. This experiment has been repeated many times and usually the failure rate is extremely high. Roughly 75% of participants give the wrong answer. Remember, the rule says *if a card has a 7 on one side, then the other side is blue*. Which cards would you flip in order to check to see whether the rule is true for these cards?

The key to solving this puzzle is realizing which cards are irrelevant to the problem. For example, turning over the blue card does not help you test the rule because the rule does not preclude cards with one blue side and a side with a number other than 7. Checking the blue card is not helpful.

Checking card #3; the card with the number 7 is important because if the card with 7 does not have a blue back then the rule is violated. Checking card #1, the card with a green side is also relevant, since if the other side had the number 7, the rule would also be violated. No relevant information can be gathered by checking cards #2 and #4. The fact that most participants in this experiment fail to correctly answer is made all the more puzzling once we realize that they almost universally understand and accept the explanation of the correct answer. If you were one of the roughly 75% of people who did not manage to solve this problem correctly, it should encourage you to continue your study of deduction in earnest.

As we will see, logic is not directly concerned with empirical questions of truth and falsity nor is it concerned with the kinds of rhetorical strategies for persuasion that occupy advertisers and politicians. Instead, logic focuses on the structure or the form of arguments. The study of logic strips away the distracting emotional or rhetorical features of arguments in order to allow us to identify and evaluate their formal features.

We can find the formal features of arguments by paying attention to the way the speaker's or author's reasoning moves from one thought or sentence to the next. These moves are **inferences** and the kind of deductive reasoning that we have been considering here moves by a sequence of inferences. After studying logic, we are in a better position to understand, evaluate, and improve the inferences that form the heart of our reasoning. Once we learn to uncover the patterns and moves that lie underneath the noisy and distracting surface rhetoric, we can evaluate arguments more easily.

> Logic focuses on the structure or form of arguments.
>
> Special emphasis is placed on the evaluation of the connections between the steps of an argument (the inferences).

The key to understanding whether an argument is formally correct involves understanding whether its conclusion **follows logically** from the basic assumptions of the argument (the premises). This idea of *following logically from* is also known as ***logical consequence.*** Understanding what it means for one sentence to follow logically from others is a large part of our project in this book.

As we have already seen, most of us have the ability to detect obvious cases where a conclusion does not follow from its premises. We are good at finding and rejecting obvious contradictions and unjustified jumps in a train of thought. However, as the Wason selection task shows, our ability, even in relatively simple cases is systematically flawed under certain circumstances. The study of logic provides a way to extend and supplement our common sense ability in complicated cases and to correct our systematic errors in simple cases.

The Wason selection task gives us evidence to believe that even in cases where there are relatively simple errors in reasoning, it can sometimes be difficult for us to spot them until we begin to engage in a little formal reasoning. The Wason task involved no distracting rhetoric or emotionally loaded terms. It was a relatively austere formal puzzle, but one that proves quite challenging most of us. Imagine how much more challenging our analysis becomes once we are faced with informal cases that involve preconceived notions and strong habitual or emotional associations.

4.5 How to Begin Formalizing an Argument

In this section, we will develop some strategies for dealing with cases where the formal error is obscured by strong associations we encounter in the language of the argument. Consider the following bad argument:

> *You should take a course that teaches you about morality and ethical decision making. If ethicists were always honest, then you could afford to be ignorant about moral questions and could simply consult the experts when you had a moral problem—but the problem is, there are dishonest ethicists, and so you must understand your own moral decision making. The best way to learn about ethics and morality is to take a course.*

The advice being given here seems, superficially to be quite reasonable. Isn't it a good thing to know something about ethics? Isn't it always prudent to be a little bit cautious when dealing with philosophers? Yes, knowing about moral philosophy is a good thing, and it is certainly a good idea to protect one's own interests in any transaction with a philosopher. **However, the argument that is given in support of the conclusion is a faulty one even if the conclusion happens to be true**. The problem becomes clear when we examine the form of the argument. At this point, let's introduce in a very rough way the kind of analysis and formalization that we can bring to arguments in order to unpack their structure and evaluate their formal features.

Treating the argument a little more formally we find that we can break the argument into parts consisting of those parts that are describing some state of affairs (declarative sentences) and the terms that connect declarative sentences (logical operators). Each of the declarative sentences is colored. The parts that serve to give the argument its logical structure are colored black.

Initial Sketch of our Faulty Argument:

If

ethicists are always honest (***P***)

then

one does not need to know about moral philosophy (***Q***)

but (and)

it is not the case that ethicists are always honest (¬***P***)

so

it is not the case that one does not need to know about moral philosophy (¬***Q***)

What we've done in this box is to take the core of the English argument in order to break it into a pair of declarative sentences; *ethicists are always honest* and *one does not need to know about moral philosophy*. These two sentences assert something about the nature of the world. Simple declarative sentences of this kind are the building blocks of logic.

The next step is to associate an italicized lowercase letter with the spot in the argument that is occupied by each declarative sentence.

"***P***" can stand in for every appearance of the declarative sentence "*ethicists are always honest*"

"***Q***" can stand in for every appearance of the declarative sentence "*one does not need to know about moral philosophy*"

In addition to the declarative sentences, this argument contains terms which perform a logical function. In this case those terms are *if. . .then*, *so*, *but*, and *not*. The logical role of the word "but" turns out to be equivalent to "and" "Not" and "not the case that" are also taken to have the same logical role in this context.

By identifying the parts of the argument that are playing a logical role and distinguishing them from those parts of the argument that make assertions about states of affairs in the world, we are already well on our way toward providing an analysis which can help us determine the legitimacy of the argument. We can go one step further by eliminating most of the English text in the argument and using variables to represent the declarative sentences as follows:

Replacing Variables for Declarative Sentences in our Faulty Argument

If P then Q

$\neg P$

Therefore, $\neg Q$

Once we have made the step of replacing declarative sentences with variables, we have the skeleton; the logical form of the argument. At this point, you might already see the problem with the argument. Sometimes, it helps to replace the variables in a formal pattern like this with simpler declarative sentences in order to understand it more clearly. It is sometimes easier to think of the same pattern in a more intuitive way using another example from natural language which takes the same form. Consider the following example which has precisely the same logical form is the case we are considering:

If I run a marathon, then I will be tired	If P then Q
I didn't run a marathon.	$\neg P$
Therefore, I'm not tired.	Therefore, not Q

Common sense immediately alerts us to the problem with this analogous piece of reasoning. After all, thinking back to Sam's faulty reasoning in Chapter 1 when he was giving advice to Karla, we recall that running marathons is not the only way that people can get tired. One can be tired because one has been studying, or playing video games all night, or because one has run a half-marathon instead of a marathon. Given the possibility that I could be tired for many different reasons, my not having run a marathon does not guarantee that I won't be tired. Therefore, it isn't necessarily the case that because I did not run a marathon, I am not tired. If there are counterexamples, of the kind we just mentioned, then the conclusion does not follow validly from the premises. If a conclusion does not follow necessarily from its premises, then it is logically implied by those premises.

Notice that the faulty marathon argument takes precisely the same form as the faulty argument about the ethicist. The ethicist argument might have *felt* more reasonable, but our feelings in that case were leading us astray. The marathon example should suffice to convince us of the fallacious nature of arguments with this form. We have yet another example of the fallacy of *denying the antecedent*.

For now, the main point is simply to introduce you to how relatively simple formalization can supplement our basic commonsense capacity to detect faulty arguments. Notice that in this case, we were initially distracted by the apparently plausible piece of advice. In order to properly evaluate the apparently reasonable argument being presented, it was necessary for us to strip away the natural language formulation of the argument. Strings of apparently plausible sentences tend to wash over us in ways that lull our critical capacities into a lazy state. If each of the sentences sounds reasonable, we are unlikely to worry too much about the structure of the argument. What formalization offers is the chance to shed the familiar associations and habitual connections that impede our critical judgment.

Exercises for Chapter 4

1. Do you believe that there are a universal set of logical norms that we should follow in reasoning? Think about the reasons that someone might doubt the existence of such norms. What would they argue should replace the norms of good reasoning?
2. Good deductive reasoning is said to be truth preserving. What does that mean and why is it important? Imagine dynamical cases in which, over the course of an argument, some sentences that were once true become false. "It is morning" is true when we begin an argument at 11.59 a.m. but false at 12.01 when we end the argument. Do cases like this have any significant implication for our view of the importance of the truth preserving nature of deductive reasoning?
3. When someone engages in fallacious reasoning, what would be the best strategy to show them their mistake? Think of someone who has never studied logic, but is a reasonably intelligent person of good will.
4. Try formalizing one of your favorite arguments. Convert the declarative sentences to variables. Leave the logical words in English.

5 Epistemic Virtues

Humans and some other animals, have the capacity to think about their social and physical surroundings and to plan actions through a process of reasoning or deliberation. We reason with varying degrees of success. Sometimes we successfully manage to make sense of our situation and can find the best course of action in light of our understanding. Sometimes we fail; we misunderstand the world, make terrible decisions, and embark on disastrous courses of action. Given the complexity of our environment and the diversity of our own preferences and values, all of us are confronted with the need to make decisions.

How can we hope to make better decisions? To begin with, it is important that we reflect on what is involved in the process of reaching a decision. In Chapters 2 and 3, we examined the basics of argument analysis and saw how good decision making is a lot like conducting an argument with oneself. We also saw how it is important to be a charitable interpreter of alternative positions and sources of evidence. In Chapter 4, we examined the nature of deductive inference and saw that there is a distinction between good and bad inference; not every move we make when we are engaged in reasoning is legitimate. Making inferences is certainly a central feature of reasoning and developing the ability to evaluate inferences is one of the important purposes of studying logic. However, in order to improve as decision makers, we need to understand the components of deliberation and we need to be able to tell when each of those parts is operating well or badly.

As we argue with ourselves about the best course of action, we engage in a wide variety of activities. A highly incomplete list of these activities would include:

- thinking through reasons for different courses of action,
- imagining alternatives,
- weighing costs and benefits,
- evaluating evidence,
- striving to honor our commitments,
- attending to our emotional responses,
- listening to what others tell us
- etc.

As we can see from this list, the deliberations that go into making a decision involve many cognitive and noncognitive resources and skills. Insofar as our decisions result from a process of deliberation, there will be some element of reasoning involved. Admittedly, some factors that shape our decisions are likely to be beyond our conscious control. But this fact should lead us to take even more care with aspects of our decision-making that do fall within our control.

Acquaintance with some formal techniques can improve the quality of our decisions if we hope to aim for excellence in reasoning. Merely getting through a logic course will not guarantee that you achieve excellence in reasoning. In fact, many excellent philosophers, logicians, and mathematicians believe foolish things and fail to make good decisions. In addition to studying the techniques of formal reasoning, it is also necessary to cultivate some of the virtues that lead to good reasoning and decision-making.

Most of us would prefer to reason or deliberate well, rather than badly. But what do we mean by "reasoning well"? As we saw in Chapter 4, logic normative questions like these are not solved by simply referring to facts about the brain or by running surveys to determine what the opinion of the majority happen to be. To reason well is to follow logical principles, to consider evidence in an unbiased manner, and to take appropriate care with our judgments with respect to important matters involving harms to ourselves and others. If we care about reasoning well, then we should aim to follow the following five rules:

Recognize the limits of your ability to reason	(Be humble)
Respect logical norms	(Be consistent)
Consider alternatives	(Be imaginative)
Change your mind when the evidence warrants it	(Be open-minded)
Don't trust your gut when it comes to probabilities	(Be statistically savvy)

These are a very rough set of principles and their associated epistemic virtues. If one is arrogant, inconsistent, unimaginative, closed-minded, and ignorant of statistics, *and* one is unwilling to overcome these problems, studying logic and other formal methods is unlikely to cultivate excellence in reasoning.

5.1 Why Aim for Consistency?

Most of us intuitively understand **consistency** to be an important virtue. In part, this is because inconsistency can hamper decision making and action. One cannot both drink one's coffee and not drink one's coffee at the same time without risking a spill. Thus, if we are interested in decisions that inform our actions, we will be interested in consistency. One of the tasks of an introductory course in logic is to help us to achieve consistency in our reasoning and in our arguments.

If we hope to avoid actions that run counter to our values and preferences we should pay attention to the quality of the reasoning that influences those decisions and strive for consistency. But what exactly is consistency? The easiest way to think about consistency and inconsistency is as a relationship between sentences. Two sentences are inconsistent if they cannot both be true at the same time:

S_1 There are more cows in Ireland than in Greenland.
S_2 There are not more cows in Ireland than in Greenland.

Together S_2 and S_1 comprise an inconsistent set of sentences. Either of these sentences could turn out to be true under certain circumstances, but you know with absolute certainty that they cannot both be true together. More generally:

> *We say that a set of sentences is inconsistent if the sentences cannot possibly all be true at the same time.*

Conversely, a set of sentences is consistent if they *can* possibly all be true at the same time.

Being consistent does not mean being stubborn or refusing to change one's mind in the face of strong countervailing evidence. In fact, refusing to revise one's views over time is likely to indicate a vice rather than a virtue. If we consider a person's development over the course of a lifetime, we can agree with the nineteenth century American philosopher Ralph Waldo Emerson when he says that "a foolish consistency is the hobgoblin of little minds". Refusing to change one's mind; refusing to modify one's beliefs in light of new evidence, would be an example of foolish consistency.

However, as we consider our own reasons for holding some view or acting in some way, we ought to be sensitive to inconsistency among our own beliefs. Inconsistency among one's beliefs at a time usually signals trouble. If my beliefs on some important matter are inconsistent then I should revisit my beliefs and find reasons for accepting one or the other of the pair of conflicting beliefs.

We saw in Chapter 1 that there are good moral reasons to value critical thinking. However, even if one is not moved by moral considerations one will be forced to take critical thinking seriously for purely selfish reasons. This is because, if we are unwilling to correct our epistemic vices then in principle we become targets for those who wish to turn us into **money pumps**. "Money pump" is a term from economic theory that is meant to describe the fate of someone who has some inconsistency in their preferences or some confusion about probability that leads them to be easily exploited.

Let's begin by considering someone who has consistent preferences. If, for example, you prefer strawberry ice cream to chocolate ice cream, you might pay a small sum to swap the chocolate for the strawberry. At the same time, you would not pay to exchange strawberry for chocolate because this would run counter to your preferences.

Imagine someone with inconsistent preferences. This is difficult to coherently imagine, but think, for example, of someone who asserts both that:

> they prefer strawberry ice cream to chocolate ice cream
>
> and
>
> that they prefer chocolate ice cream to strawberry ice cream.

When asked if they would like to pay a small amount to swap strawberry for chocolate they would accept the deal. When asked subsequently if they would like to pay a small amount to swap chocolate for strawberry they would also accept the deal. After these two payments are made they would return to their original position of having strawberry ice cream, but would be poorer as a result of the two exchanges. The person who discovered their inconsistency has converted them into a money pump and can drain them of cash until they correct their vulnerability in some way. In such a simple case, it is genuinely difficult to imagine a typical adult human not quickly realizing what is happening and making a rapid adjustment.

Reflexive, Symmetrical, and Transitive Relations

In addition to being transitive or intransitive, binary relations can have the property of being symmetrical or asymmetrical, and reflexive or irreflexive.

Transitive Binary Relations:

If we consider some set of objects S we say that a binary relation R is transitive for that set S if

> for any three members of S
>
> x, y, z
>
> if xRy and yRz,
>
> then xRz.

For example "bigger than" is a transitive relation

Reflexive Binary Relations:

R is reflexive if for all *x*

xRx.

An example of a reflexive relation would be "equal to" or "same height as".

Symmetric Binary Relations:

R is symmetric if for all *x,y,*

if *xRy,* then *yRx*

An example of a symmetric relation would be "sibling of."

Economists were not concerned with obvious inconsistencies of this kind. Instead, they were interested in more complex cases where the trouble is less obvious. Specifically, they were concerned with cases where an agent's preferences are not **transitive**. Transitivity is a property of some relations. For example, the relation of being *taller than* is transitive whereas the relation *being in love with* is not. Here's how it works: when we say that a relation like being *taller than* is transitive what we mean is that

if

a is taller than *b*

and

b is taller than *c*,

then

a is taller than *c.*

By contrast, the relation of being *in love with* is not transitive meaning that we cannot assume that because *a* is in love with *b* and if *b* is in love with *c*, therefore *a* is in love with *c.*

Returning now to our preferences: Imagine that I am presented with some ice cream choices and I have some set of preferences with respect to flavors. Here's what I like:

I prefer chocolate to orange.

I prefer orange to banana.

I prefer banana to chocolate.

Let's first notice that my preferences are not transitive. If my preferences were transitive, then when I prefer chocolate to orange and orange to banana, I should prefer chocolate to banana. My ranking of preferred flavors should be well-ordered if I am to avoid becoming a money pump. In our example, they are not. This might be excusable in some sufficiently long list of options where one could easily imagine losing track of what one prefers to what. If I failed to notice this for some reason, then my preferences can be exploited. Specifically, if I stuck to my preferences and didn't notice what was happening to my financial situation, an unscrupulous ice cream dealer can use three flavors of ice cream to drain my bank account. Here's how:

> Ice cream seller: "John, here's some orange ice cream, but I also have chocolate, it's just a penny to swap, would you prefer that?"
>
> John: "Yes, I prefer chocolate to orange and am willing to spend a penny to satisfy that preference, so here's a penny."
>
> Ice cream seller: "Great, but hold on I also have banana, it's just a penny to swap, would you prefer that?"
>
> John: "Yes, I prefer banana to chocolate and am willing to spend a penny to satisfy that preference, so here's a penny."
>
> Ice cream seller: "Great, but hold on I also have orange, it's just a penny to swap, would you prefer that?"
>
> John: "Yes, I prefer orange to banana and am willing to spend a penny to satisfy that preference, so here's a penny."

At this point, I am back where I started with orange ice cream (perhaps already melting) and three pennies poorer. Since I'm a savvy customer and can tell that there is financial trouble ahead for me if this continues, I will refuse his offer of an exchange for chocolate. I can see where our transactions are leading—even though I *really do* prefer chocolate to orange. If we were to find ourselves in situations like this, we would adjust our behavior in order to avoid becoming bankrupt.

In economics, it is thought that some trade or series of trades in which one of the participants is always worse off than the other is very rare and should not be a persistent feature of markets. This is because one party will realize, as John does in the example above, that the irrational features of his behavior and preferences are harming him.

The phenomenon of intransitive preference is only one way in which irrational behavior can result in our being exploited. It is an instance of what economists call a **Dutch Book**. A Dutch Book is a scenario in which the casino sets odds on a game of chance such that the set-up guarantees a profit for the house, regardless of the outcome of the gamble. In order to be turned into a money pump, you need only to be irrational about the way that probabilities work and to be unlucky enough to have that irrationality exploited.

Most people recognize that they are limited in a variety of ways. When it comes to reasoning and decision making, it is important to recognize that we are susceptible to making a wide range of errors. If we reflect on an occasion where we were mistaken about some important matter, we can ask whether our error was an isolated case, or whether we might also be mistaken with respect to other important beliefs? Only a seriously deluded person will fail to recognize that they have held false beliefs, made poor decisions, or have reasoned badly in the past. Epistemic humility begins with the recognition that we are not always right about our beliefs and that there is room for us to improve as thinkers. Once we accept that this is the case, we have reason to begin cultivating the other epistemic virtues.

5.2 Imagination

In addition to wanting to be consistent, most of us would prefer to avoid being unimaginative thinkers. If we are unable to ask good questions or imagine alternatives, it is likely that our decision making will be less effective than we would prefer. Unimaginative thinkers are likely to miss opportunities and to remain stuck in unsuccessful habits of thought and action.

Part of what it means to reason well is to have a well-developed imagination. One useful byproduct of one's study of logic, as we shall see, is the cultivation of one's imagination. The simplest way to develop one's imagination is to begin with an exercise like the following:

> Begin by simply reflecting carefully on any of the many beliefs you take to be true.
>
> The next step is to consider the denial of that belief.
>
> Now, consider how things would be if that belief were false.
>
> In case it were false, what else would be different about the world?

Imagination can be prompted into action simply by asking "What if. . .?" What if I'm wrong about there being no life on the moon? Almost immediately, I am imagining the kinds of creatures that might be living on the moon. If you are stuck for things to imagine, simply consider one of your ordinary beliefs and ask "What if I'm wrong about that belief?" Simply considering the possibility that you might be wrong about some belief is the first step toward cultivating your theoretical imagination.

> What if I'm wrong about the importance of privacy?
>
> What if democracy is not the best form of government?

> Aristotle noted that one of the marks of an educated mind is the ability to entertain a thought without accepting it.

Notice that entertaining the thought that democracy might not be the best form of government, or that privacy might be an overrated concept does not commit you to being against democracy or privacy. Instead, it is likely that by carefully considering the alternatives, your understanding of political questions will improve. You will be a more competent advocate for democracy, for example, if you have carefully considered the alternatives and can explain why you believe that democracy should be preferred.

5.3 Be Ready to Change Your Mind When the Evidence Warrants It

As we shall see, conclusions reached via valid deductive inferences follow necessarily. What that means is that in the case of deductively valid inferences, if the assumptions or premises are true, then the conclusion *must* be true. Being able to achieve certainty about some matter is one of the excellent features of formal reasoning.

However, most of the inferences we make in ordinary life and in scientific inquiry do not have the kind of absolute certainty we sometimes find in logic and other formal domains. Instead, generally we must settle for reasoning that gives us pretty good, but not necessarily certain, conclusions. If we can be reasonably confident that some conclusion follows from the evidence, this is usually satisfactory. Personal experience, the evidence of our best science, and trusted testimony are the imperfect sources that we rely upon in both scientific and ordinary reasoning. Furthermore, when we are forced to make decisions, we often do not have the luxury of perfectly reliable information.

Many of the most important generalizations we find in the sciences are discovered via experience, controlled experiments, or computer modeling. Any of these methods admits of error and therefore good scientists are willing to revise their commitments in light of compelling evidence to the contrary. Good scientific practice involves recognizing the possibility that one is wrong about ones claims. Even if this possibility is very small, **a scientist must be willing to admit the possibility that her beliefs about the natural world are subject to revision**.

> When trying to determine the truth in most matters, we can usually do no better than rely on the best scientific evidence that is available to us, while knowing full well that scientific claims are subject to error.
>
> While we are often unable to achieve certainty, we would be foolish to opt for anything other than what we regard as the view that is most likely, or probably true.

In ordinary life and scientific practice, we frequently employ what are called ***inductive inferences***. Imagine walking through a bookstore for example in noticing that the titles of the books close to the entrance are in English. It would be reasonable to expect, based on this evidence, that, as an English speaker in an English-speaking country, you would be able to read most of the books in the store. Inductive inference works in a similar way; we take some observed pattern or sample of evidence and from this we infer the probable truth of some conclusion. Thus, when I walked into the bookstore I observed a pattern, namely that all the books in the sample that I saw were in English. I concluded from this evidence that the books I would encounter later in the store would probably also be in English. In making this inference, I am assuming that the pattern I noticed initially would continue throughout the store. It is possible for me to be surprised, and to discover, for instance the bookstore has a sizable foreign language section. However, given my initial evidence and my experience with bookstores in the past, my inference was reasonable.

Compare the following patterns of deductive and inductive reasoning:

> Deductive:
>
> All the books in this store are in English.
> *Wuthering Heights* is in this store
> Therefore, *Wuthering Heights* is in English.

> Inductive:
>
> The bookstores in this town mostly carry English-language books.
> A book called *Akira* is on sale in a bookstore in this town.
> Therefore, *Akira* is in English.

In the case of deductive inference, we usually say that **an argument is deductively valid in case it would be impossible for premises to be true and the conclusion false**. Deductive inference will be a central topic for us going forward. Inductive inferences do not have the same kind of necessity attached to them. For an inductive inference to be good, we simply need the premises to help us see that the conclusion is *probably* true. In the inductive argument above, it might be the case that the premises are true while the conclusion is false. It might actually be the case for example that a copy of *Akira* we find in our local bookstore is in Japanese. The fact that this book turned out to be in Japanese does not mean that it is false to say that the bookstores in this town mostly carry English-language books.

The kind of inductive reasoning, which we considered above, is an instance of extrapolating or projecting from a sample of some objects of interest to the whole set of those objects. There are no guarantees that we will always see the same pattern in the relationship between the premises and the conclusions of an inductive inference. For example, the fact that the value of equities in the stock market has tended to increase over time is no guarantee that their values will continue to increase: Past performance is no guarantee of future results.

The fact that I've never become ill after eating at my favorite restaurant is no guarantee that it won't happen one of these days. And yet, as the great Scottish philosopher David Hume noted, while it's true that inductive judgments provide no guarantees and should not be regarded as having anything close to the certainty of deductive judgments, reliance on induction is an unavoidable part of human life. The fact that I cannot be absolutely certain that my sandwich will not poison me does not stop me from eating my sandwich. As another great philosopher, John Locke pointed out, a person "that in the ordinary Affairs of Life, would admit of nothing but direct plain Demonstration, would be sure of nothing, in this World, but of perishing quickly." (E IV.xi.10: 636) What he meant by this was that we inevitably encounter some degree of uncertainty when it comes to our ordinary decision making.

Inductive reasoning involves reasoning with uncertainty. As I sit down to lunch, I reason that the sandwich is probably not going to poison me. The probability that the sandwich will poison me is slim, but it is not zero and before I eat the sandwich I must decide whether the effort involved in making sure that the sandwich is not poisoned is too great given the very slim chance of real danger. I cannot rule out the possibility that this sandwich will kill me, but I judge from past experience that it probably won't and I am willing to take the risk.

A hated dictator will be more cautious about his sandwiches than a relatively anonymous philosopher or plumber needs to be. But none of us will be absolutely sure that we will survive our next meal. It might be the case that the dictator employs sandwich tasters to make sure that his lunch is safe to eat. But even this won't guarantee his safety. We can't count on inductive inferences with certainty, but we are stuck having to make them.

5.4 Think Carefully about Risks and Probabilities

How careful we need to be influences how strict we are with respect to inductive judgments. Our view of the risks involved in some course of action are closely connected with the standards by which we judge inductive inferences. For example, in cases where we don't have any real reason to care, we are likely to be relatively lax in our standards. Our inductive judgments are treated with more strictness and attention when the costs of being wrong are more severe. For example, most of the time you probably trust your doctor's judgment. However, in cases where the health concern is more important, or where the decision involves greater cost, it is prudent to seek a second or third opinion. If the doctor advises you to have your leg amputated, the price of incorrect advice is high. By contrast, if the doctor simply writes you a prescription for antibiotics for a nasty cold you are unlikely to question the doctor's judgment.

Reasoning about probability can become quite difficult and we will devote some attention in later chapters to some of the technical details. For now, let's think about a case where common sense and reasoning about probability are intertwined. As you consider the four scenarios below, try to keep track of your own beliefs. Try to see whether your views and feelings about risk change as you read:

A. It's late at night on a camping trip and as you sit around the campfire, talking late into the night your good friend who is normally reliable and trustworthy tells you a detailed and terrifying story about how he was abducted by aliens. The story takes the familiar form of others in the genre, paralysis, tractor beams, grey-skinned aliens with large black oval eyes, examination tables, probes, and the like. He tells the story in a sincere and heartfelt way. Two weeks later, that same friend asks you to give him some money so that he can cover his house in aluminum foil. How much money are you willing to give him?

B. It's late at night at home and on your TV you see an advertisement for an elaborate electronic hat-like device that, the expert on the TV (with a PhD after his name) promise offers protection from alien abduction. The hat costs less than $10 shipping and handling included. How cheap would the hat need to be in order for you to consider buying it? Would you wear one to bed if it cost nothing to get one?

C. Because of widespread concern in the population, your nation's government decides to impose a modest tax increase on its citizens in order to build an elaborate network of satellites to intercept and destroy alien spacecraft in order to protect the population against the threat of alien abduction. The impact of the tax on you will be about $50 per year. How do you view the tax?

D. You are listening to a talk-radio program featuring Dr. Herman Himmel a Professor in Southwest Northern University. He is described by the host as a highly-regarded parapsychologist. He has just published a book claiming that abductions are not actually being conducted by extraterrestrials. Instead, he claims to have evidence that shadow people from the center of the Earth are responsible. He is giving his book away for free on his webpage. Will you take the time to read it?

Consider your reasoning in each of these scenarios. Does your reasoning about risk shift as you move from case to case?

It is highly unlikely that aliens are abducting people. However, you might be more likely to believe that aliens are abducting people if a close friend or an allegedly authoritative figure tells you that this is the case. In each case, you must decide whether to put your time or resources into the prevention of alien abduction. If you are relatively confident that aliens are abducting people and consider the experience a serious unpleasantness to be avoided, then you should be willing to invest resources in preventing the abduction. How much you are willing to invest will indicate something about your estimation of the probability of an abduction actually happening and the cost of reducing the risk. If you were certain that the aliens were coming for you, you would likely spend far more on protection against aliens than you are currently spending.

We are not in possession of detailed knowledge of the probabilities involved in whether aliens are really abducting people. Common sense can serve us well in cases of this kind. An advanced civilization whose members are capable of traveling across vast expanses of space is not likely to need to abduct humans for science experiments or other purposes. Common sense should tell us that their limited abilities in biology would be difficult to reconcile with their excellent understanding of physics. Note also their apparent predilection for American rather than say Chinese, Brazilian, or Swedish victims. Are the aliens somehow uninterested in the biology of Swedes, or is there some other explanation for the large number of American victims? Notice too that the widespread availability of mobile phone technology seems to have reduced the occurrence of UFO and alien phenomena in recent years.

Common sense encourages us to reject the idea of alien abduction, but it cannot eliminate all possible worries. It is possible that these aliens actually do invest the necessary resources in visiting and abducting Americans. Since we are not absolutely certain that aliens are not going to abduct us, the challenge is to determine how much you would be willing to pay to insure against the small probability that such an unpleasant encounter would befall you.

In a less exotic case, how much would you pay to insure against having your car destroyed by a meteorite? Cars have been hit by meteorites in the past and it is possible that it could happen to you. Your answer should depend on the value of your car, on the cost of the insurance, and your estimation of the risk. In Chapter 9, we will examine the formal aspects of thinking about risks and probabilities in detail.

In a much more ordinary case, how much should we pay to avoid putting our children in danger in the event of a car accident? For example, many parents in richer parts of the world strap their children into car seats to mitigate the risk of harm during an accident. In North America or Europe, the cost of a car seat is widely considered a reasonable price to pay to reduce this risk. The prospect of losing one's child in a car accident is horrible, but it is a risk that we cannot eliminate entirely without foregoing cars completely and thereby assuming other kinds of risk. Is there a cost beyond which we would not ask parents to pay? If the safest car seat on the market cost $5000, should we force parents to borrow money to cover the cost? At what point is the cost of reducing risk not worth it?

Before you insist that you would pay any price to protect your children, remember that money spent on reducing that risk is money which cannot be spent on education, food, medicine, toys, piano lessons, and so on. Given finite resources we must be sensitive to the fact that decisions involve trade-offs. The term "**opportunity cost**" is used to mark the fact that decisions involve costs. Here's why:

> Making decisions involves foregoing mutually exclusive alternative decisions. If I decide to do *A* I often thereby decide not to do some other action, call it *B*. Perhaps doing *B* has some benefit associated with it. I will have to sacrifice this benefit by choosing *A*. The benefit that I am foregoing is the *opportunity cost* of choosing *A*. In deciding to spend one's time or money in pursuit of some good, one is also deciding to not spend it in some other way.
>
> In our example, if I decide to buy an expensive car seat for my child, I am not spending my money on other things that would benefit my child. The marginal improvement in my child's safety

> with an expensive rather than an ordinary car seat might not be worth the opportunity cost. In general, we must evaluate whether spending time or money in pursuit of some good outcome might not be worth the opportunity costs incurred thereby. Maybe saving for my child's education is a better use of that extra money than buying the slightly safer car seat.

Critical thinkers make decisions with an eye to opportunities that will be foregone. When we make decisions that have significant consequences for the well-being of others, it is especially important that that we recognize our duty to avoid negligent decision making. Note that in our example, the parent who spent money on an extremely expensive car seat would be blameworthy even though he or she acted on praiseworthy motivations, namely out of love for his or her child and out of concern for his or her child's safety.

The fact that a decision is motivated by high levels of empathy, love, loyalty, patriotism, or other praiseworthy feelings does not always make it good. Imagine a highly empathic father who uses force to prevent his child from being hurt by another person. Without further details, we would likely praise the imagined father. However, there are scenarios, for example, in the context of a childhood vaccination, where his use of force against a nurse administering painful injections to his child would be blameworthy in spite of his empathic motivation.

While we might praise the father for his love of his child, he is blameworthy insofar as he is failing to exhibit the epistemic virtues. Lack of epistemic virtue leads to unethical behavior, even by nice people.

Exercises for Chapter 5

1. It is often said that extraordinary claims require extraordinary evidence. Consider a conspiracy theory like, for example, the claim that the moon landings were a hoax. Why is this an extraordinary claim and what kind of evidence would serve to convince you of its truth?
2. What is homeopathic medicine? Why do some people use it on themselves and their families? Given that homeopathic medicine is basically just water, is there any reason to discourage people from purchasing it?
3. Consider a situation in which you were required to think carefully about opportunity cost. How did you make your decision? How did you decide what the opportunity cost was?
4. Cultivating imagination is an important part of critical thinking. One way to stimulate one's imagination is to deny some deeply held belief. Once one has done this, consider how the world would be different in this alternative reality where one's deeply held belief is false.
5. People with very different political views from you are very like you in most ways. Why do so many of us regard those with different ideological or moral perspectives as monsters or idiots? Try to imagine a good, intelligent person holding ideological views that are at odds with yours.

6 Bias, Heuristics, and Argument Patterns

How can logic and other formal methods help us in our decision making? Surely, the decisions facing us are highly particular, sensitive to context, with their own characteristic set of benefits and costs? It is true that many of the most important decisions we face are complicated, emotionally sensitive, and involve incomplete information. Our decisions are also influenced by the natural forces that shaped our mental capacities. As embodied, biological creatures our minds are at least partly the product of the evolutionary history of our species. Perfection is an expensive luxury and our minds evolved reliable, but imperfect solutions to the problems faced by our ancestors. Finite resources and adaptive pressures led to the development of cognitive short-cuts and rules of thumb that can predictably trip us up. In order to understand where we are liable to go wrong, we must know something about the psychology of reasoning and decision making. Thus, a good course in critical thinking will draw on insights from psychology and economics, in addition to philosophy and the purely formal sciences.

> If the pattern, or form of the argument is bad, then the argument itself is bad and should be rejected.

In general, our goal is to detect and evaluate patterns that recur in our own reasoning and decision making and in the thinking of others. Recognizing patterns of reasoning provides an efficient way to begin evaluating specific arguments. Some of these argument patterns are good, others bad. We have already seen some obvious ways that argument patterns can be bad and in the chapters that follow we will learn some techniques for evaluating arguments in greater detail. The reason that his is an efficient strategy is that if our reasoning fits a pattern that we know to be bad, as soon as we detect the pattern, we can reject the argument without being concerned with the details or the specific content of the argument itself.

Chapter 2 introduced the idea of an argument and noted that while we usually think of arguments or discussions as involving two or more persons, one can also privately engage in an argument with oneself. We considered the deliberations of a student who attempted to decide whether to continue with their education and we began to develop our ability to distinguish reasons for and against choosing one course of action over another. The purpose of this chapter is to study patterns that repeatedly appear in ordinary reasoning with special attention to the kinds of biases that psychologists and economists have shown affect our reasoning and decision making. As we shall see, even when we understand that a pattern of argument is a bad one, we are often inclined to commit the error because of innate tendencies in human reasoning.

6.1 Considering Values as We Deliberate

As one considers the right course of action to take, one might begin thinking through reasons for opting for one course of action over another. But how do we know what to aim for? What kinds of outcomes are we hoping for? In modern life, we are often encouraged to think of our decisions in quasi-financial terms. We are often told that we ought to compare the costs and benefits of one course of action over another. In one respect, this is a very sensible approach. However, by itself, this accounting model is an insufficient guide. To see why, it is worth asking whether deliberating carefully is simply a matter of calculating costs and benefits. Can't we just compare the list of pros and the list of cons separately and compare the two lists as we saw the college student attempting to do in Chapter 2? Unfortunately, the task is not that simple. To begin with, in order to even being measuring costs and benefits we must answer some difficult questions:

What values are we trying to maximize?

> For example, even if I am a particularly venal character and am motivated solely by self-interest. Am I solely interested in getting as much pleasure as possible or as much money as possible? I might notice that maximizing money might conflict with maximizing happiness or pleasure. Earning money takes time away from other pleasures and a life devoted to the consumption of opiates might maximize the experience of physical pleasure (at least for a while), while draining one's bank account. Thus, even the hedonist has to weigh competing values.

Do I hope to honor other commitments or achieve other kinds of good in addition to personal pleasure?

> For example, my pleasure or happiness might not be my primary reason for pursuing some course of action. I might be interested in maximizing the well-being of others, or following some moral principle unrelated to pleasure I might be motivated by the pursuit of excellence in some field and might be willing to subordinate personal happiness or well-being for that achievement.

How do we establish a clear standard for measuring the values of distinct outcomes?

> Perhaps there are a variety of important values that I am committed to. Maybe I am interested in both my own happiness, the well-being of my children, and in the pursuit of excellence in my chosen career. How do I balance the three motivating values?

Even if we opt for a cost-benefit approach to making a decision we should begin by settling some of these questions. Decision making is clearly more difficult than a simple arithmetical problem insofar as we need some clear sense of what we care about and what outcomes we are aiming to achieve. In spite of the difficulties, if we hope to become good decision makers, we must judge the merits of the arguments in favor of different courses of action. We should aim to have the best reasons available as we make our decisions.

The question of what we care about, and more importantly, the challenges related to knowing what we ought to care about, are beyond the scope of this book. Nevertheless, it is important to notice that good critical thinker must be thoughtful about the goals and values that orient decision making. It is not enough to weigh costs and benefits without thinking seriously about what one takes to be a cost and a benefit or what one takes to be a bad and good outcome. Good critical thinkers are not blind to the moral complexity of decision making, nor are they uncritical about their own preferences. This is especially true for those of us who live in a cultural environment in which our preferences are subject to manipulation by powerful commercial and political agents.

Once we have a sense for what we value and what ends we are hoping to pursue, the easiest way to begin tackling the diversity and complexity of decision making is to learn to detect patterns of good and bad reasoning. Chapter 4 introduced some simple examples of good patterns of inference (like modus ponens) and bad patterns of inference (like the fallacy of denying the antecedent). When it comes to decision making, we can also detect patterns of reasoning that are easy to identify and reappear regularly in our own experience. We will begin with a very simple, but very common argument form. Our goal is to understand how it figures in our thinking, what is wrong with it, and how we can identify other instances of the same form.

Our goal in the next section will be to understand the problem with this argument pattern but just as importantly, we should work to understand the psychological factors that encourage us to repeat this faulty pattern of reasoning. This argument pattern is known as “the argument from sunk cost” and the psychological tendency that is connected to it is a very common cognitive bias, known as loss aversion.

6.2 The Argument from Sunk Cost

Let’s imagine the following internal monologue that a homeowner might have with himself:

Should I renovate the kitchen of this old house?

> *I'm not sure whether a new kitchen will make any difference to the price of the house when we're ready to sell and I hate cooking anyway. . .*
>
> *But, I've invested so much time, money and energy renovating the rest of this place, it would be a shame not to finish. . .*
>
> *OK, I'll start working on it.*

What is the context of this little internal monologue? Clearly, they have to make some decision concerning how she will expend their time and resources. They consider some reasons in favor of and against renovating the kitchen. Then they arrive at some conclusion based on what they regard as the reasons in favor of renovating the kitchen.

Perhaps you or I will never face a decision exactly like this one. However, even if one never owns an old house, or never has to decide on renovations, one can recognize the homeowner's reasoning as following a familiar pattern. Economists and philosophers call this pattern of reasoning the *argument from sunk costs*. The pattern illegitimately motivates many of our decisions. Philosophers sometimes call patterns like these *argumentation schemes*[1]. In this particular argumentation scheme, we are claiming that if we do not continue some course of action, we will be "wasting" the time and resources that we have already spent. The trouble with this way of thinking is that the time and energy that we are worried about wasting have already been spent and are not coming back. From the incorrect concern that we have with waste we move to the conclusion that we should continue the course of action that we are presently on.

Arguing from sunk costs is a mistake for the following reasons. As one decides on future renovations, the fact that one has already fixed the bathroom and living room are not relevant to and should not influence one's evaluation of whether fixing the kitchen is worth the expense. Notice that it might be worth fixing the kitchen but *the reasons for doing so are independent of the costs that have already been spent on other things*. The reasons given in the argument above "I've

[1] Walton, Douglas, Christopher Reed, and Fabrizio Macagno. *Argumentation schemes*. Cambridge University Press, 2008.

invested so much time, money and energy renovating the rest of this place, it would be a shame not to finish" do not, in fact, support the conclusion.

There are good reasons for renovating a kitchen:

If one spends a great deal of time in the kitchen, then it makes sense to make it a pleasant place to be.

Similarly,

If a good kitchen increases the resale value of the house more than the cost of renovating the kitchen, then this fact would also serve as a very good reason to renovate.

However, the fact that you have used resources to renovate the house in the past, is not, by itself, a good reason to continue investing resources in the project.

> The bare bones of the argument from sunk costs look like this:
>
> Question: Should I continue this course of action?
>
> Premise: I have already invested a great deal of my resources in this course of action
>
> Conclusion: Therefore I should continue this course of action

The argument from sunk costs is repeated in a variety of contexts and it is often a highly persuasive rhetorical strategy. Consider the following example of an argument from sunk cost:

"Happyland lost 21,435 members of our army in *Operation Destroy the Evil Ones*, I would like to remind the advocates of withdrawal that we owe our dead something. . .We will finish the task that they gave their lives for. We will honor their sacrifice by staying on the offensive against the Evil Ones and building strong allies in the Northern Territories that will help us fight and win the war on evil."

"I know you're unhappy and our relationship is not good, but we've been together for three years. You can't break up with me after all we've put into this."

> "This slice of cheesecake is too big and I'm feeling full, but I shouldn't waste it, it cost me $8."
>
> "We just heard that there's a great party over at Mike's house. But we can't go because we have these tickets for a boring play."
>
> "I hear the outlook for this company is very grim and the stock price is likely to fall further, the trouble is, I've already lost so much money on it (I bought it at $10 and now it's down to $3). I shouldn't sell now, maybe it'll recover."

Each of the examples above takes the form of the sunk cost fallacy. If we recognize that the same pattern of reasoning is present in the case of the kitchen renovation and in the case of the argument for "staying the course" in a military engagement, we can adopt a more sophisticated critical stance toward arguments of this kind. Sometimes, strong emotional forces might cloud our judgment. However, our judgment as to whether we should agree to continue to send people into a war should be based on the merits of doing so, rather than on the fact that we have already incurred great costs.

Unfortunately we will be unable to recover past losses of life and treasure no matter what course of action we decide upon at present. The dead are dead and the resources that we have spent on war cannot be recouped. These facts are independent of the decision facing the government of Happyland. The decision in question concerned the merit of expending *new* lives and *additional* treasure in the future.

If this is your first time thinking about the sunk-cost fallacy, you are likely to be having two conflicting sentiments. On the one hand, it is completely obvious that just because we *have* spent money, time, or effort on some project does not mean that we should *continue to* spend money, time, or effort on it. Strictly speaking, the merits or pitfalls of an investment are independent of our past decisions. However, on the other hand, the sunk cost fallacy is seductive because we have a deep-seated tendency that makes us feel as though we should not "waste" the past investment of energy or money even in cases like the ones we considered above. Later in this chapter we will examine why strong psychological dispositions like these play such an important (and often damaging) role in our decision making.

After considering the argument from sunk costs, consider your feelings as you think through the following example:

> Imagine that you would really like to own a Gibson Les Paul guitar, but since they are too expensive, you opt for a cheaper guitar instead. The cheaper guitar, the Stratozapper, costs \$1000 but you only have enough money to give the seller a deposit for \$500 while you work to save enough money to complete the transaction. After a few weeks, you have finally saved the other \$500 when you see a special offer at the local guitar store: A Gibson Les Paul is on sale for \$500. What do you do?

Since you have been thinking about sunk costs, you are now likely to make the correct decision. However, for many of us it is difficult to recognize that the decision as to what to do with the \$500 in your pocket is independent of the fact that you have already left a \$500 deposit on the less attractive guitar. Many of us tend to think something like: "I've already invested \$500 in the Stratozapper, it would be a shame to waste that money." We can now see that this line of thinking is irrational. It is easy to see the problem when we can represent the situation as follows:

Decision point →

(a) [Lose the deposit] and spend \$500 on the Les Paul.

(b) [Lose the deposit] and spend \$500 on a guitar you would prefer less than the Les Paul.

In both cases, you would have spent the \$500 and lost the \$500 deposit. The difference is that in (a) you get the guitar you prefer, whereas in (b) you get a suboptimal guitar. The rational choice is (a). But notice that for most of us there is still some sense that not opting for (b) means "wasting" the initial deposit.

Part of the task of logic and formal reasoning more generally is to allow us to see patterns of argumentation across diverse cases in order to give us a way of easily detecting successful arguments and common errors. Now that we have studied the sunk cost fallacy, we will be able to recognize instances of this argument pattern in novel cases and will be more able to resist the temptation to act in irrational ways because of them.

Our tendency to fall prey to the sunk cost fallacy is explained by a deeply rooted tendency in human psychology called **loss aversion**. *Loss aversion is our tendency to have an unbalanced concerned for losses over gains.* The problem for rational decision making is that because we more strongly prefer to not lose something than we would prefer to gain something of equal value, we tend to misunderstand the actual costs and rewards of course of action.

It seems to be a simple fact about our psychology that our happiness is reduced more by losing \$100 than it would be increased after winning \$100.[2] A strong concern with losing is a feature of our psychology that undoubtedly served us well in our evolutionary history where holding on to current resources, rather than taking a risk for future gain was probably important for survival value. However, it can also lead us to systematically bad patterns of reasoning. Imagine being offered the following opportunity:

> Given a fair coin, I will give you \$2 for heads, but if the coin comes up tails, you will give me \$1. Rationally, you should take the opportunity since the coin toss has an expected return of \$.50 ([\$2 × 0.5] – [\$1 × 0.5])
>
> But what if the cost of losing were \$100 and the prize for winning were \$200?[3]

[2] Kahneman, Daniel, and Amos Tversky. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the econometric society* (1979): 263–291.

[3] Benartzi, Shlomo, and Richard H. Thaler. "Myopic loss aversion and the equity premium puzzle." *The Quarterly Journal of Economics* 110.1 (1995): 73–92.

Understanding how we feel about losses like this is a complex matter. Perhaps I would take the first bet and shy away from the second because losing $100 would be so harmful to me that I would forego what I recognize to be a great deal. The second offer is just as good as the first in terms of the odds, but I might feel freer to make the first bet than the second because, for example, not having $100 might lead to catastrophe whereas not having $1 is not such a big deal for me. The problem with loss aversion is that it encourages us to overestimate how bad it is to lose and leads us to miss out on possible opportunities.

Understanding that we can have an exaggerated aversion to losses in ways that can systematically disadvantage us is relevant to making decisions in a range of contexts. Not only are we subject to loss aversion in financial decisions but also, for example, in the decision as to whether to end a romantic relationship, leave a job, move to another city, and so on.

How might loss aversion influence one's thinking as one decides whether to leave an unhappy relationship or an unsatisfying job? How might loss aversion prevent us from taking advantage of potentially fruitful opportunities?

In thinking about these questions, it is useful to think about what psychologists and complexity scientists call the **explore/exploit** conflict. Living things face the problem of deciding between a conservative and an adventurous strategy. The conservative strategy or what is more commonly called the *exploit* strategy recommends sticking with where we are and what we are doing in pursuit of known and familiar rewards. The adventurous, or *explore* strategy recommends trying less familiar or well-known options or alternatives in search of better rewards than are currently available. The trade-off between exploring and exploiting is easy to see. On the one hand a conservative strategy, the apparently sure-bet risks foregoing potentially rich rewards that might lie just beyond the agent's horizon.[4] In a competitive context, the conservative strategy also risks allowing more courageous competitors to flourish, perhaps returning home to devour or defeat their conservative adversary.

[4] Wilson, Robert C., et al. "Humans use directed and random exploration to solve the explore–exploit dilemma." *Journal of Experimental Psychology: General* 143.6 (2014): 2074.

As we have seen in the case of loss aversion, we humans tend (in general) to favor the conservative/exploit strategy over the adventurous/explore strategy. In some important contexts, of the kinds mentioned above, this tendency leads us to make errors. In the next section, we will examine tendencies of this kind in more detail.

6.3 Heuristics

There are some facts about our limitations as thinkers that we cannot avoid. For example, as we saw in the previous section, we are subject to loss aversion in ways that leave our reasoning vulnerable to the sunk cost fallacy. We are embodied beings in the physical world and as a result our cognitive capacities are limited. Our ability to remember, for example, is limited such that our ability to follow a chain of reasoning becomes gradually less reliable as arguments become longer or more complicated.[5]

Insofar as we human beings are rational, our rationality is limited. In the 1950s, Herbert Simon coined the phrase ***bounded rationality*** to describe the way that human reasoning is constrained by incomplete information and finite resources.[6] Simon's insights led psychologists and economists to take the limits of our reasoning ability seriously rather than assuming that human decision makers behaved like ideal rational agents.

Our reasoning is not only limited by having finite resources, it is also biased. These biases come into play even when we have access to complete information about a problem. Loss aversion, for example, is a bias on our reasoning that is not due to finite resources. Instead, loss aversion is something like a hard-wired or innate disposition toward a particular strategy of reasoning. This section explores some of these innate biases.

In the late 1960s and early 1970s, psychologist Amos Tversky and economist Daniel Kahneman were working together to understand

[5] Miller, George A. "The magical number seven, plus or minus two: some limits on our capacity for processing information." *Psychological Review* 63.2 (1956): 81.

[6] Simon, Herbert Alexander. *Models of bounded rationality: Empirically grounded economic reason*. Vol. 3. MIT press, 1982.

the challenges faced by people solving problems involving judgments concerning statistics and probability. They had discovered that even trained mathematicians had trouble reliably applying their knowledge to some of the tasks they designed. As they studied the way that people make mistakes, Tversky and Kahneman started to detect some regular features of human reasoning that they called *cognitive biases*. By **cognitive bias,** they meant our shared tendency to fail to make rational inferences in certain kinds of reasoning.[7]

In recent decades, following the work of Tversky and Kahneman, psychologists identified dozens of similar patterns of faulty reasoning. Cognitive biases are associated with psychological tendencies to reason using simple shortcuts or heuristics. Other cognitive fallacies can be ascribed to limits on our cognitive capacities.

Let's explore some of the biases that characterize human thinking in order to develop some strategies for counteracting their negative effects. Some of these biases seem to be innate features of our minds, shared by most members of our species. Some biases appear to be culturally specific, reflecting social conditions in different places and times. Most obviously, for example, biases that inform judgments concerning class, gender, and race will vary depending on historical and cultural contexts. Illegitimate social or cultural biases not only lead to some of us being treated unfairly, they also result in less successful and less well-informed decision making. If we hope to achieve excellence in reasoning, it will be important to understand the effect of bias no matter what its source. If, for example, we pay less attention to the opinions of someone from Mississippi because we have the incorrect belief that people with southern accents are less intelligent than average, we are making a mistake in reasoning that results from a socially conditioned bias.

The kinds of cognitive biases that Tversky and Kahneman discovered do not appear to be socially conditioned. Instead, they result from the action of innate **cognitive heuristics** in our reasoning. A cognitive heuristic is something like a rule of thumb, or a tendency

[7] Kahneman, Daniel, and Amos Tversky. "Subjective probability: A judgment of representativeness." *The concept of probability in psychological experiments*. Springer Netherlands, 1972. 25–48.

that we employ in our thinking (without really thinking about the fact that we're using them). Psychologists explain many of these heuristics as adaptive. What they mean by describing heuristics as "adaptive" is simply that they evolved as a way of making our thinking more efficient and improving the survival prospects of our ancestors. Generally, these heuristics operate below the level of conscious thinking and allow us to make snap decisions without taking precious time for extended deliberation.

In the context within which our species evolved, understanding of complicated statistical questions, for example, had little relevance to survival. Such questions were certainly of less relevance to our survival than the ability to recognize obvious associations quickly and efficiently. Tversky and Kahneman (1974, 1124) introduce the role of heuristics in our reasoning by comparison in the following way: Consider a rule that might apply in our judgments about things we observe in our visual experience. We might have a rough rule of thumb like the following:

The more sharply an object is seen, the closer you should judge it to be.

Notice that, strictly speaking, this advice is based on a false generalization about vision and distance. Most of the time, following this rule of thumb works well as we attempt to judge how near or far the objects in our visual field are from us. The rule is not a guide to the truth in every case, but it is reasonably reliable under ordinary circumstances. It is easy to imagine the kinds of exceptions to the rule that would lead to error in a system that followed it rigidly. These exceptions mean that automatically following the rule would cause us to make systematic errors in our judgment of distance in cases where visibility was either poor or unusually good. We could predict that in conditions where the air is clean and humidity is low, systems that obeyed this heuristic would, at least at first, incorrectly judge objects to be closer than would be the case.

While the heuristic would systematically fail in cases like those considered above, it might be the case that systems that follow this rule could be more efficient and faster over the long run and for the most part than systems that follow more complicated and demanding rules. One reason why a rule like this is conducive to success might be the fact that it is simple. An animal or a system that relied on this rule would, in general spend less time on judgments about distance, could react more quickly, and therefore would be likely to be more successful and efficient than its competition. Nevertheless, this heuristic introduces a bias into judgments and can prove dangerous in some cases. Our evolutionary history has generated a situation in which the risk of occasional errors is outweighed by the benefits of a system that is generally efficient and fast.

> A heuristic does not have to be perfect in order to be adaptive. In fact, if being perfect takes time and resources from other important aspects of the organism's business, then the faster, cheaper solution is likely to persist through evolutionary history.

6.3.1 The Representativeness Heuristic

The representativeness heuristic is our tendency to misjudge the probability that, for example, not knowing anything about the patient other than the fact that he or she has a torn ACL***, we are likely to overestimate the odds that the person is a football player based on strong association we have between football players and torn ACLs.*** In Chapter 9, we will examine this heuristic in more detail and will examine some of occasions where we make systematic errors in reasoning because of it.

6.4 Cognitive Biases

6.4.1 The Availability Heuristic and the Power of Anecdotes

Judgments concerning probabilities are often challenging and complicated for us. However, in our evolutionary past, our ancestors faced uncertainty and were forced to make judgments, for example, concerning the behavior of friends, enemies, or predators. We humans evolved some general strategies for judging risks. Prominent in this context is the **availability heuristic**. This heuristic causes us to favor solutions or answers that come easily to mind over those that are less easily remembered.

When we try to judge the likelihood of an event we often rely on the ease with which we can imagine or remember similar events. For example, if one is asked to judge the probability of suffering a stroke or getting in a car accident one is likely to begin by scanning one's memory for occasions where things like this have happened to friends or acquaintances. If I know many people who have been in car accidents, or if I have been in a car accident recently, I am likely to judge the probability of getting in a car accident to be relatively high. In general, this heuristic makes good sense. If things happen often, then one is likely to more easily remember them. If it is easier to remember them then they are more readily available. Thus, generally speaking, it was adaptive for human beings to use availability as a mark of high probability.

However, as Tversky and Kahneman pointed out, availability is not due solely to the frequency of an event. There are many reasons that some idea or memory might be available to us over and above the number of times that the event or object in question has been encountered. Many of our most powerful memories are of events that happened only once. A single event with a strong emotional component may be much more available to us than boring events that happen very frequently. We tend to forget the vast number of uneventful car rides or plane trips we might have taken whereas a bad accident will remain prominent and vivid in our memories.

If the availability or retrievability of an event were simply a matter of how frequently we encountered that event then the availability heuristic would be a reasonably reliable (but again, not a perfectly reliable) guide to probability in one's immediate environment. But of course, the frequency of something happening in one's own experience is not generalizable to the experiences of others. For example, I lived many years in Boston, and as a cyclist, I learned to expect rude and aggressive behavior from drivers. However, my memories of the driving habits of Bostonians do not serve me well on the polite and friendly streets of my current home Lawrence, Kansas.

The fact that one's individual experiences and memories are often not generalizable is only one, relatively minor reason to beware of the misleading effects of the availability heuristic. As Tversky and Kahneman showed, a variety of factors in addition to the frequency of an event influence its availability or retrievability. The classic experiment they conducted to detect the effect of the availability heuristic involved giving subjects a list of names consisting of men and women of varying degrees of fame. Afterward, subjects were asked whether the list they had heard contained more male or female names.

In some lists, the men were relatively more famous than the women and in others, the women were relatively more famous than the men. In each case, subjects erroneously judged the lists with the more famous men to have more men in it. Similarly, in lists with the more famous women in it, subjects erroneously judged the list to contain more women than men (1973, 11). The more famous names were more available to the subjects and the ease with which subjects recalled the names overshadowed their memories of the numbers of men and women in each list.

The availability heuristic leads to a range of mistakes in decision making. To understand one way in which it might work, think of the role of **anecdotal evidence** in your own thinking. An anecdote is simply a story (usually a short story) that is often offered as evidence for some claim. Consider the following three anecdotes:

> *My uncle was robbed when he visited Costa Rica.*
>
> *Darmoon had a bad cold, but after he ran five miles it went away.*
>
> *Jessica had really bad luck with identity theft until she uninstalled all the apps on her phone.*

Each of these anecdotes can be understood to serve as reasons for you to believe something. Either that Costa Rica is dangerous, that running cures colds, or that downloading apps onto your phone makes you vulnerable to identity theft. I hope that at this point in the book, you are hesitating before taking any of these anecdotes too seriously as significant evidence. Is a single event really sufficient to underwrite a scientific generalization? At this point, we need to think about what counts as genuine evidence. Does the fact that one's uncle was robbed in Costa Rica warrant the judgment that it is a dangerous country?

What if one's uncle was the only person robbed in the country in forty years? We could ask similar questions about the other two examples. Anecdotal evidence is usually not very good evidence. To see how the power of anecdotal evidence is drawn from the availability bias, imagine the following scenario:

> You are making a decision to buy a particular brand of computer. You are likely to do your research on the merits of this kind of computer relative to others. Perhaps you have reviewed some reputable tech websites and have studied their reviews of the computers. You might have read the rankings provided by *Consumer Reports* an unbiased, independent, nonprofit organization that tests products using good scientific methods. Let's assume that these studies involve a large sample of machines and are conducted in a scientifically legitimate manner. You are leaning towards buying the top-rated model from Company X, but are also interested in the runner-up computer, the Company Y product. You make your decision, pick up the box and walk to the checkout. Your phone rings and it is your friend Stephanie from New Zealand.
>
> "So what are you up to?"
>
> "Actually, I'm just buying a computer"
>
> "Cool, what kind are you getting?"
>
> "Oh, it's a Company X brand laptop"
>
> "Oh, man, my friend bought a Company X laptop and he's had all kinds of problems with it"
>
> "Really?"
>
> "Yeah"
>
> "Hmmm, maybe I should get the Company Y laptop instead"

In changing your mind about your choice, you have ignored all of your previous research instead swayed by Stephanie's story. Anecdotal evidence is of significantly less evidential value than the studies conducted by reliable scientific methods. After all, her anecdote concerns just one instance of the product. Presumably if their studies were properly conducted, they would have already taken into account the fact that sometimes machines of all kinds break. Even the most reliable products can sometimes have faulty instances. However, anecdotal evidence is difficult to resist. After all, while you do not know Stephanie's friend you do know Stephanie and her story has a kind of vividness that the spreadsheet in a report on a webpage or a page of data in a magazine might not have.

The availability heuristic is operating in your decision when you are motivated by Stephanie's anecdote to buy Computer Y instead of X. It is much easier for us to call to mind a story told by our friend Stephanie than for us to think back to the studies and reviews that we had read earlier in the day. However, the anecdote is a single datum whereas the studies and reviews you had relied upon were the result of extensive testing.

The availability heuristic biases our judgments in many ways. Think of your judgment of the danger of dying in a terrorist attack or being assaulted by a stranger. You might have access to reliable statistics on the likelihood of unpleasant events like dying in a terrorist attack and you might know, on some level, that your chances of dying in a terrorist attack are negligibly small. Nevertheless, given the powerful emotional associations and memories that many of us have, it is difficult to rationally judge the appropriate level of resources to devote to worrying about terrorism as opposed to for example reducing the prevalence of heart disease, diabetes, or suicide. I might understand very well that I am far more likely to die of heart disease than from a terrorist attack and yet still fail to correctly prioritize terrorism in the list of genuinely dangerous things in my life.

6.4.2 Anchoring

Another very striking phenomenon, known as the **anchoring effect** involves a peculiar feature of reasoning wherein initial information, which may or may not have anything directly to do with the topic under consideration influences our judgment in a way which leads to biased outcomes. The best way of introducing this phenomenon is to consider the example of prices in a car showroom. The price that is listed on the sticker serves as an anchor for the negotiation to follow.

Whether we believe that we have gotten a good deal or a bad deal is influenced by the anchor. Naturally, in cases like this the anchor is set to deliberately influence negotiations in favor of a seller. When shopping for a car, we should not assume that the sticker price should serve as the anchor for the reference point for negotiation. However, it is difficult to completely insulate our negotiation with the salesperson from the anchoring effect of the sticker price. In this section, we will explore why it is that anchors play such a powerful role in our decision making.

In another classic experiment, Tversky and Kahneman asked their subjects to estimate the percentage of African countries in the United Nations. Each subject spun a Wheel of Fortune with numbers between zero and 100. The wheel was rigged so that it stopped at either 10 or 65. After spinning the wheel, subjects were asked whether the percentage of African countries in the United Nations was higher or lower than the value they were given by the wheel of Fortune. Most people who receive 10 on the wheel of Fortune would say that more than 10% of the countries in the United Nations are African countries while most people who receive 65 as a value in the wheel of Fortune would say that less than 65% of the countries in the United Nations are African countries. So far so good.

Next, they were asked to estimate what the percentage of African countries actually is. Strikingly, their estimates depended on whether they had received 10 or 65 from the wheel of Fortune. Rationally speaking, there should be absolutely no relationship between one's estimate of the number of African countries and the results of spinning the wheel of Fortune. However, the effect was profound. The group who were given 10 as the result on the wheel of Fortune gave as their median estimate that 25% of countries in the United Nations were African countries whereas the group who were given 65 as the result on the wheel of Fortune gave a median estimate of 45%.

These experiments, and others like it, show that an arbitrary starting point like the results given by a wheel of fortune significantly influenced the way that subjects estimated a completely unrelated value. Similar effects have been demonstrated using other completely arbitrary values. This powerful result has been demonstrated repeatedly and shows how careful we need to be with respect to anchoring effects of various kinds.

It is now a well-established result in psychology that our estimates can be influenced by asking us to consider an entirely uninformative number. Understanding exactly why the anchoring effect is so powerful is a topic of ongoing research. However, in a wide range of contexts, from estimating the value of a house to negotiating salary we should be sensitive to the role that anchors play in our reasoning.

It is very difficult to avoid the effects of anchoring, even in cases where we are aware of the phenomenon and understand that we are subject to its influence. In a recent book (2011) Daniel Kahneman suggests that the effect is so powerful that the best strategy in response to the effect of anchoring is to simply step away from any important decision where anchoring is in force. In practice, for example, he recommends simply walking away from a negotiation where one's competitor sets an unreasonably high or low number as the starting point. His reasons for suggesting this strategy are simple. In contexts where anchoring effects may play a role there is a huge advantage to being the first player to make a move. Establishing an anchor inevitably influences negotiations, even in cases where both parties realize that the initial number on the table is entirely unrealistic. Similarly, in cases where we plan to make large purchases; buying a car or house, for example, it is important to set one's own anchor point to the extent it is possible to do so. When buying a car for example it is worth knowing precisely how much you are willing to pay for a car and then in the negotiation with the salesperson it is prudent to ignore the anchors that he or she offers. This is a difficult practice, however reminding oneself of the existence of the anchoring effect is a useful first step in avoiding its pitfalls.

6.4.3 Confirmation Bias

Effective reasoning and decision making requires that we take evidence seriously. Our understanding of reality should be sensitive to evidence in such a way as to maximize the fit between the world and our beliefs about the world. Good critical thinkers are open to changing their minds and, as we have seen, they revise their beliefs when the evidence is sufficient to do so. However, understanding what counts as good evidence is complicated by our biases. As we try to evaluate evidence or interpret data we tend to favor evidence that supports our cherished beliefs (positive evidence for our beliefs) and we tend to ignore or discount evidence against those beliefs. When we interpret evidence in this way, we are falling prey to what is called the **confirmation bias**.

In our daily lives, we gravitate toward sources of information that come from the kinds of people that we would like to affiliate with. We are entitled, of course, to associate with whomever we prefer. Trouble arises when we gather our information solely from people who agree with us. In the United States, news sources tend to attract people who identify particular socio-economic or political groups. Readers of *The New York Times* or *The Washington Post* both of which are generally more aligned with the ideology of the Democratic Party are more likely to want to associate with each other than they are with readers of the Republican leaning *Washington Times*. More starkly, people who enjoy watching Fox News are likely to characterize themselves in opposition to people who enjoy listening to National Public Radio and vice versa. The trouble with gleaning one's news from sources that already basically agree with one's political perspective is that we inevitably fall prey to confirmation bias.

In social media the so-called echo chamber effect is amplified. If, for example, one's Facebook news feed is populated by the posts from one's friends, and if one's friends share similar values and perspectives, then it is likely that users feel their own prejudices and perspectives are being reinforced and confirmed. In fact, one's Facebook news feed provides a narrow window on the world. It is shaped first by a select handful of friends and second by algorithms that are constructed to select those posts from friends that are most likely to be clicked by the user.

Rather than hearing only from people whose perspectives are like our own, critical thinkers should hope to hear from those who will disconfirm our favored views. Rather than getting their evidence through a narrow window of like-minded sources, a critical thinker will seek out the strongest opposing voices to see whether their own views can stand the test of serious scrutiny.

The term "confirmation bias" was introduced into the psychological literature by Peter Wason in the early 1960s.[8] Wason was one of

[8] Wason, Peter C. "On the failure to eliminate hypotheses in a conceptual task." *Quarterly Journal of Experimental Psychology* 12.3 (1960): 129–140.

the pioneering figures in the psychology of reasoning and we have encountered his famous selection task already in Chapter 3. Recall that in those experiments, subjects had difficulty reasoning correctly about simple patterns of inference. Wason not only showed how difficult it is for us to reason about simple formal patterns, he also showed how difficult it is for us to appropriately evaluate evidence.

Here is his classic experiment: Wason presented subjects with the sequence of numbers 2-4-6 and then asked them to identify the rule that governed the sequence. At this point, it is worth considering your own reaction. What do you think the rule governing this sequence is? Now, how would you go about confirming whether your hypothesis is correct? Wason's experiment gave participants the following method: In order to discover whether their hypothesis was correct they could propose additional sequences of numbers and the experimenter would respond by telling the subjects whether those sequences followed the rule or not. When most of us hear the sequence 2-4-6, we think that the rule governing the sequence is "add two to the previous number" or "list an ascending sequence of even numbers". We are likely to propose tests like "8-10-12", "6-8-10", "100-102-104", and so on . What subjects were not generally inclined to do was to suggest sequences that might falsify their hypothesis. They sought evidence that confirmed their hypothesis rather than disconfirming it. Wason's actual rule was not "list an ascending sequence of even numbers" rather it was simply, "list an ascending sequence of numbers." Thus, 1-2-3 and 5-30-156 counts as obeying the rule, whereas 3-2-1 or 455-555-222 do not. Discovering the genuine rule involves seeking evidence that would help to rule out alternative hypotheses rather than seeking evidence that supports one's existing guess. Wason's original experiments have been subject to considerable debate and varying interpretations. A range of explanations have been offered to account for the fact that we generally tend to favor evidence that supports our favored hypotheses and that we neglect evidence that does not support our hypothesis.[9] In the intervening decades, confirmation bias has come to be widely

[9] For a detailed discussion of the history of the experimental study of confirmation bias see Nickerson, Raymond S. "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of General Psychology 2.2* (1998): 175.

recognized as a general tendency in human reasoning that leads us to favor some evidence over others solely because it tends to support a position that we prefer.

As it is commonly understood, confirmation bias leads to a selective consideration of evidence such that we reinforce our existing beliefs and neglect evidence which might contradict those beliefs. The more deeply held or entrenched those beliefs are the stronger the confirmation bias generally turns out to be. For example, in studying arguments for or against emotionally-charged topics like abortion, gun control, or the legalization of narcotics, people will systematically favor evidence, which supports their view and will discount evidence which might place their view in doubt. Obviously, these topics are more emotionally charged in the United States and less so elsewhere. In contexts where the emotional stakes are not as high, we would see less effect from the confirmation bias.

It might seem relatively obvious that people would be biased in their evaluation of evidence with respect to deeply held or emotionally sensitive claims. The trouble with confirmation bias is that we rarely recognize its influence on our own beliefs. It is easy to say that other people are selective about evidence or that they fail to pay attention to the views of the other side. However, confirmation bias is also in play in your own thinking. As we saw above, few of us actively seek out news on current events from sources that hold political positions that are contrary to our own. For example, few of us tend to read blogs or opinion pieces that we know ahead of time are likely to criticize positions we hold dear. Overcoming confirmation bias requires actively considering the possibility of error and perhaps even actively seeking out arguments that are opposed to our own positions. It is very easy, thanks to the Internet, to seek out sources that confirm our own views and to customize our consumption of media in such a way as to insulate ourselves from countervailing opinions and arguments.

Obviously, we must be selective about the sources of information and opinion that we consume. However, in order to avoid having our decisions undermined by confirmation bias it is worth including some sources of news and information from a variety of perspectives.

In cases where one has to make an important decision, it is worth asking whether the manner by which we arrived at our reasons was affected by confirmation bias. Deliberately and carefully considering counterarguments or counterexamples to the principal reasons informing your decisions is one way to combat this bias.

Checking for Confirmation Bias:

Pause

Why am I making this decision?

Because of x

What is my evidence that x?

> Can I be certain that I haven't overlooked some evidence that would show that x is false?
> How confident should I be that x?

However, confirmation bias is not simply a matter of people being irrationally prejudiced by their emotional state. Even in situations where are emotional commitments are weak and tempers are cool we tend to search for information in ways which confirm our preferred beliefs. In part, this results from our tendency to prefer hearing and providing positive responses to questions than negative ones.

One effect of this tendency can be found in opinion polling. For example, depending on the phrasing of a question, we will get very different sets of results depending on whether the response is elicited in a negative or a positive way. The following questions are asking about the same set of facts:

> Are you satisfied with your current employer?
>
> Are you unsatisfied with your current employer?

If you answered positively to the first question then if you are consistent, you should answer negatively to the second question. If you are satisfied with your current employer you should not be unsatisfied with your current employer. However, people who are asked the first question systematically respond in a way that reflects more positively on their current employer than people who are asked the second question. The proportion of "yes" responses on the first question is larger than the proportion of "no" responses on the second question.[10]

Why should this be? It appears that, in part, people simply prefer saying "yes" to "no". There are a range of complex issues at play here and this is a topic of ongoing research, however, it is clear that whether a question is phrased in a way that elicits a positive rather than a negative response given the same facts, leads to different responses.

[10] Kunda, Ziva; Fong, G.T.; Sanitoso, R.; Reber, E. (1993), "Directional questions direct self-conceptions". *Journal of Experimental Social Psychology (Society of Experimental Social Psychology)* 29: 62–63.

Exercises for Chapter 6

1. We are not required to be consistent in order to make decisions. We can reason in an illogical or contradictory way while still making decisions, what then is the relationship (if any) between logic and decision making?

2. According to philosophers like Frege, all scientific explanations presuppose that scientists accept some basic principles of logic. On this view, since scientists have to try to avoid contradictions and must adhere to the laws of logic in their own work, it would be a mistake to claim that their sciences can explain logic. He argued that this was because any such explanations would themselves presuppose logic. Other philosophers argue that we must assume some basic logical framework in order to have meaningful disagreements. Do you think that progress in science or even the possibility of disagreement between scientists requires that we agree on a common logic beforehand?

3. Logic is a normative discipline. In what ways, if any, is it similar to ethics?

4. Give some examples of faulty reasoning from your own experience. Why are these examples of faulty reasoning? Identify other cases of faulty reasoning which go wrong for the same reasons.

5. Find an editorial from a newspaper. Try to identify its main thesis. Does the editorial make appeals to the emotions of the readers? In what ways? When is it legitimate to make appeals to emotion in an argument?

6. "If fortune tellers could tell the future, they would not be telling the future for relatively small sums of money." How might someone arrive at this conclusion?

7. "I would do anything to protect my child against harm." Evaluate this claim.

8. "Only the claims of science are to be accepted." Evaluate this claim? Can the person making this claim believe it?

7 Soundness and Validity

In the previous chapter, we saw a range of psychological factors that can impede our ability to reason well. However, we did not explain precisely what we mean by calling some pattern of reasoning good or bad. In this chapter, we will introduce two important standards for judging the merits of claims and arguments: Truth and formal correctness.

We are often told by pundits and some politicians that we live in a post-truth or post-fact era. What they mean to say, of course, is that respect for the norms of honesty and rationality have declined dramatically in contemporary culture and politics in the United States and other Western countries.

There are many plausible explanations for why this decline has taken place. Part of the blame lies with purveyors of bad philosophy and their followers in the humanities. During the 1980s and 1990s, many American and European academics in the softer parts of the humanities and social sciences argued that the concept of truth is an old-fashioned relic of the enlightenment. According to the more radical among them, there simply is no truth. The more palatable, consumer-friendly version of this radical position is the notion that each of us, or at least each culture "has their own truth." Some suggested that truth and objectivity more broadly, are masculine and Eurocentric concepts that have oppressive effects on people who are nonmasculine or non-European. By the end of the Twentieth Century a well-meaning politics of toleration and anti-imperialism combined with philosophical confusion encouraged relativism about the concept of truth. Relativism is the view that there is no objectively true or false claim. Instead, judgments of truth and falsity are relative to particular conventions or cultural contexts. This perspective found a receptive audience in consumerist cultures of North America and Western Europe.

In recent decades, declining commitment to the norms of truthfulness and rationality in cultural life opened the door for politicians to come to power through shamelessly crafting false, but attractive narratives and denying the right of any authority to correct those narratives. Relativism about the truth allows political figures to talk of "alternative facts" and allows ideologues on the left and right to ignore scientific evidence that contradicts their cherished beliefs.

The relativist's assertion; "Truth is relative", is self-undermining insofar as it denies the very possibility of making a true assertion.

Is "Truth is relative" objectively true?

If it is, then at least one sentence is objectively true and relativism about truth does not hold for all truths.

If it is not, then why should it be believed?

The denial of truth is not only self-undermining, but it is also politically dangerous insofar as it allows powerful agents to avoid accountability. We live in an era when citizens are confused about whether there can be any reliable or unbiased source of information, when they are encouraged to believe that all sources of information have the same standing—that newspapers like *The National Inquirer* are just as reliable as *The New York Times*. When basic norms of rationality and honesty are not followed, rascals can hide behind the idea that truth or other values are relative. Perhaps the most pernicious consequence of relativism is that it allows wrongdoers to deny that there are any real standards by which their actions can be judged.

7.1 Truth and Formal Correctness

For our purposes, an argument can be judged by

the truth of its sentences, (TRUTH)

and

its validity (FORMAL CORRECTNESS)

For the most part, logic cannot help us to decide whether a sentence is factually true in the standard sense of telling us, for example, whether it is raining in Beijing right now, the population of Seattle on January 1, 1990, the number of times Brazil has won the World Cup, or the number of angels that can stand on the head of a pin.

Usually, but not always, the best way to discover whether some sentence is true or false would be to follow the method of rational inquiry that we find exemplified in the natural sciences. While science does not answer all questions, including some very important questions about science itself, it provides a much better model for successful inquiry onto most topics than its competitors.

Hylas: "We should only believe claims that are supported by science"

Philonius: What about the claim that you just made Hylas?

Logic, by itself, does not answer the usual questions we might have concerning the way the actual world happens to be. However, logic does allow us access to an infinite number of ***logical truths***. Logical truths are sentences that are true not just in the actual world, but in all possible ways things could be. As we shall see, there are an infinite number of logical truths of varying degrees of complexity. Our first example of a logical truth is one of the simplest we could possibly have:

There is a largest number or there is not a largest number.

You need no mathematical training to recognize that this sentence is true ***by virtue of its form***. We use the term ***tautology*** to refer to sentences like this that are true by virtue of their logical form. Let's unpack the form of our example:

let "*A*" stand for "There is a largest number"

and

"not *A*" stand for "There is not a largest number"

Notice the role of the word "or" connecting the left and right side of the sentence:

[There is a largest number] or [there is not a largest number].

A or not *A*

Common sense tells us that either "*A*" or the denial of "*A*" must be true. If this is the case, then we can conclude that the sentence as a whole must be true. If we are right about this, then all sentences that have the form

p or not *p*

(where *p* stands for any declarative sentence)

are true.

Here, we are encountering one of the basic principles in logic; **the law of excluded middle.** The law of excluded middle says that a declarative sentence and the denial of a sentence cannot be legitimately asserted together. For example, I cannot legitimately assert that there is a largest number and there is not a largest number.

More formally,

It is not the case that "*p* and not *p*."

If we accept that, notice that we must also accept "*p* or not *p*." To put it another way, the principle of excluded middle says that either a declarative sentence is true or the denial of that declarative sentence is true, there is no third option.

At this point, it might have occurred to you that there are some sentences that are always false by virtue of their form; **contradictions**. Compare our example of a tautology with the following contradiction:

There is a largest number and there is not a largest number.

The only difference between our contradiction and our tautology is the "and" and the "or."

Returning to our example "There is a largest number or there is not a largest number," we can experiment by replacing the parts of the sentence flanking the "or" with other examples. By thinking carefully about examples, we should be able to convince ourselves that any sentence of the form

q or not q

is true no matter what. Later we will demonstrate this more conclusively, but at this point we will rely on common sense. Here, the letter "q" is serving as a **variable**. As we have already seen, in sentential logic, a variable is simply serving as a placeholder for any **declarative sentence**. In algebra, letters function as variables standing in for numbers. In the early stages of learning logic, we are interested in the relationships between declarative sentences rather than numbers. In later chapters, we will see other roles that variables can play in logic.

As we will see below, **a declarative sentence is a sentence that makes a claim about the way the world is**. For the most part, declarative sentences are either true or false.

Paradoxical Sentences

For the purposes of the logic you will study in this book, you will not need to concern yourself with declarative sentences that are neither true nor false. We will touch on paradoxical sentences because they are philosophically interesting, but they are an advanced topic that will have to wait until we have the fundamentals in place.

The sentence underneath this sentence in this box is false

The sentence above this sentence in this box is true

Setting aside the challenges posed by paradox, our study of logic begins by restricting ourselves to the relationships between normal declarative sentences. We use variables to stand for declarative sentences with clear truth values as a way of drawing our attention to the formal features of patterns of reasoning. Replacing declarative sentences with variables allows us to focus on patterns of reasoning rather than being distracted by the meaning of the declarative sentences. Variables allow us to begin to abstract away from the content of the sentences in order to see structure more clearly.

The best way to begin developing our ability to reason involves paying attention to the forms of good and bad arguments. The reasons why we should focus on the form of argument are straightforward: Formal logic allows us to avoid the distraction of emotions, vague language, and strong habitual associations in arguments by focusing only on the structure of arguments. As we have seen in Chapter 6, our ability to reason correctly is susceptible to all kinds of negative influences from our biases.

Emotions, prejudice, vagueness, habits, and heuristics have their place in ordinary life and are not necessarily bad in themselves. However, they can easily impede us in the pursuit of excellent reasoning. The exercise of eliminating most of the nonlogical features of an argument and looking as closely as possible at the form of an argument is a useful antidote to these distractions. Much of the interest of formal logic is that it allows us to study the forms of arguments without having to worry about their content or meaning. Reflecting on the form of an argument rather than its content is one easy way of protecting our reasoning from distractions.

Logicians cannot always tell us whether a statement is true or false, but they can often tell us whether arguments have the right form. One of our first tasks, as we study logic, is to become familiar with this idea of the form of arguments.

The first step, as we have seen above, involves separating our usual concern for truth and falsity from our concern with the form of arguments. For example, let's consider an argument composed of a string of true sentences:

> NASA decommissioned the Space Shuttle.
>
> Why? Because Neil Armstrong was an astronaut, and Sally Ride was an astronaut,

This is an obviously bad argument. Nevertheless, it is a bad argument that contains true sentences. It is true that NASA decommissioned the Space Shuttle and it is true that Neil Armstrong and Sally

Ride were astronauts. It is not true that there is any explanatory relationship between these three facts. The problem with this argument is a problem of form. While there are many ways for an argument to be good, the special focus of our study in this book is the *formal features* of argument. The formal strengths and weaknesses of arguments will concern us here. As we have already seen, many arguments fail because of their form. Let's explore what it means for an argument to be formally correct and incorrect. Here is a formally correct argument built of false sentences:

Paris is south of Turin

Turin is south of Cairo

Therefore,

Paris is south of Cairo

Since the reasoning in this argument is formally correct, if the premises of this argument had been true, then the conclusion would have to be true. As noted above, it is a mark of formally correct reasoning that it cannot take us from true premises to false conclusions. However, the formal correctness of an argument cannot guarantee that the premises of an argument are true.

7.2 Distinguishing Soundness and Validity

Philosophers use the term "**valid**" to indicate that an argument is formally correct. The first point to notice is that validity is purely a matter of the form of arguments. A valid argument is one whose form is such that it is impossible for the premises of the argument to be true and the conclusion of the argument to be false. **In a valid argument if the premises are true, the conclusion must be true.**

By contrast, we say that **an argument is sound, if it has true premises and is valid**. It is important to distinguish soundness and validity. An unsound argument is one that contains false premises or one that is invalid. An invalid argument can have true premises and conclusions. Note that we can, for example, have invalid arguments whose conclusions are true and valid arguments whose premises are false.

Valid arguments are such that if the premises are true, the conclusion must be true.

An argument is sound if and only if it has true premises and is valid.

Let's consider some of the examples of sound and valid arguments

Sound:

All states have capital cities (true)
Kansas is a state (true)
Therefore, Kansas has a capital city (true)

Valid but not sound:

All states have monuments to Darth Vader (false)
Wichita is a state (false)
Therefore, Wichita has a monument to Darth Vader (false)

From a logical point of view, a failure in the form of argument is simply devastating. As we shall see, this kind of failure dooms an argument even if the argument is highly persuasive, and even if this defective argument is presented in support of a contention that happens to be true. In this book, the formal characteristics of argument will occupy us more than any other.

What is the form of an argument? At their most basic, as we have already seen, arguments generally consist of a conclusion, some initial assumptions or facts that are assumed (premises), and the steps that take one from these premises to the conclusion. Steps or moves in an argument are either **warranted** or unwarranted. In the case of a purely logical inference, warrant is relatively easy to determine. *A logical inference is warranted insofar as it follows the rules of logical inference*. We will explain and justify the rules of inference later in the book. Other, more informal kinds of warrant are not so straightforward.

In the kinds of formal settings that will serve as the focus of later chapters, an argument's conclusion is simply the final step in a chain of reasoning. By contrast, in informal settings, the conclusion can appear at any point in the argument and may even be difficult to detect. In previous chapters, we examined some guidelines for finding the conclusion or main contention of an argument. Once we determine the conclusion, we are in a position to evaluate the chain of reasoning that supports it.

As we examine the internal structure of an argument, we will search for the sequences of inferences that are either explicitly or implicitly supporting the conclusion. At this stage, our task is to begin the process of evaluating the correctness of these steps.

In order to find inferences, it will be necessary to concern ourselves with the way that certain basic logical terms behave in the sentences that constitute the argument. In English, terms and phrases like "because of . . . ," "and," "or," not," "all," "some," "if . . . then," "therefore," "hence," and many others, indicate logically interesting connections between sentences or parts of sentences. These terms mark the formal structure of the argument and are subject to rules of inference in ways that we will discuss in detail later.

For now, it is enough to recognize that the truth and falsity of statements should be carefully distinguished from the form of a sequence of statements.

Separating Soundness and Validity

As we begin to separate the form of arguments from the truth value of the statements in an argument, we notice that a valid argument with false premises can have a false conclusion. We notice also that an invalid argument can lead to a true conclusion. Stranger still, as we shall see later, in standard first-order logic, contradictory or *necessarily false premises* can validly imply a true conclusion!

Yes, according to classical logic, we can legitimately derive true statements from premises which cannot be true.

7.3 Validity and Counterexamples

Validity is a formal property of arguments; some arguments are valid, some are not. We say that an argument is valid if and only if its conclusion is true whenever its premises are true. **The relation between the premises and the conclusion of a valid argument is known as "entailment."** For sentence, C to be entailed by another sentence P means that C logically follows from P. An argument is valid if its conclusion is entailed by, or logically follows from its premises. What it means for one statement to logically *follow from* or be *entailed by* other sentences is a challenging philosophical and technical problem and it will occupy is in more detail in later chapters. For now, we will repeat the most important feature of an argument's validity: **If an argument is valid, then we know that if the premises of the argument are true, the conclusion *must* also be true.**

As we shall see, there are cases where a false conclusion can follow validly from false premises and where a true conclusion can follow validly from false premises. There are also cases where the conclusion of an invalid argument is true. However, **it is *never* the case that false conclusions follow validly from true premises**.

We determine the validity of an argument by checking its form. As we shall see, **validity is entirely a matter of the form of the argument**. As we have seen, a formally correct or valid argument can be composed of false or nonsensical sentences. When we examine validity in more detail, we will concentrate almost entirely on form.

What do we mean by the form of an argument? For starters, let's look at a few bad arguments, all of which fail by virtue of their formal features. We will begin with less obviously bad arguments, before looking at more obviously bad arguments. What we shall see is that these bad arguments all share the same formal structure. The purpose of looking at these bad arguments is to understand how to recognize their formal characteristics. The following is an example of bad reasoning:

(1)

Premise

Jorge is opposed to labor unions, nationalizing health care, and government-funded public education.

Premise

Followers of Ayn Rand are opposed to labor unions, nationalized health care, and government-funded public education.

Conclusion

So, he must be a follower of Ayn Rand.

You might think that there is something plausible about this line of reasoning. After all, he opposes unions, public education, and nationalized healthcare in a manner strongly reminiscent of followers of the philosopher Ayn Rand and therefore, it might seem reasonable to infer from this that he is may well be a fan of Ayn Rand. What we know about his political views might support the belief that he is highly likely to have an interest in Rand and given a great deal of additional context and information about probabilities, this might be a reasonable assumption. However, as it stands, apart from other extra information, the argument itself is faulty. Think, for example, of the claim that he *must* be a follower of Ayn Rand. Surely, it's possible that he might hold these opinions and never have heard of Ayn Rand. Alternatively, he might be an anarchist who happens to be opposed to labor unions or any other kind of political organization.

Denying that he is a follower of Ayn Rand does not contradict any of the premises of this argument. Given what the premises tell us, it might not be the case that he is a follower of Ayn Rand. If the conclusion followed logically from the premises, then we would know with absolute certainty, given truth of the premises that he is a follower of Ayn Rand.

The formal failure of the argument about Jorge, as it stands, becomes clearer when we look at analogous and slightly more obvious cases of bad reasoning. Notice that the argument about Jorge takes the same form as the following example of bad reasoning:

(2)

With the exception of 2, all prime numbers are odd.
15 is odd,
Therefore, 15 is a prime number.

A prime number is a number which is only divisible by itself and 1. Since 15 is not prime, something has gone wrong in this piece of reasoning. Someone making an argument like this one is misled by the association of the properties of being odd and the properties of being prime. However, even though all prime numbers (other than 2) are odd, not all odd numbers are prime. Just as with the case of Jorge, we are misled by the strong association between two characteristics. (In that case, being a fan of Ayn Rand's philosophy and opposing nationalized healthcare, labor unions, and government-funded education. In this case, being odd and being prime.)

Consider a third analogous case, which should make the failure of the first two even more obvious:

(3)

All dentists are human.
Sally is a human.
Therefore, Sally is a dentist.

Commonsense tells us that this is obviously an intuitively unacceptable inference. However, it has the same formal structure as the other cases. It's easy to see that the third example is problematic. Since this bad argument takes the same form as the other two, we should be suspicious of them too. In the third case, we can see that the claim that Sally is a dentist is not supported by the fact that she is a human and the fact that dentists are human. After all, it is fully compatible with her humanity and with the humanity of all dentists, that she is an airline pilot. We will examine errors of reasoning like this in more detail later. For now, we just need to recognize that the form of arguments is important to good and bad reasoning.

We can see, just using our common sense that (3) is a bad piece of reasoning. Since (1) and (2) have the same form, they are also bad pieces of reasoning. We latch onto strong associations in cases (1) and (2), and these can sometimes keep us from seeing the problem as easily as we can see it in the case of (3).

The fact that there are **counterexamples** to an argument means that the argument is invalid. In a valid argument, there are no counterexamples. In this context, a counterexample is simply an exception to the claim that something *must* be the case. What is a counterexample precisely? Very simply, it is a way that things can be that denies some claim. For example, the claim that all Irish people are alcoholics is defeated by the existence of a nonalcoholic Irish person. The first nonalcoholic Irish man or woman we meet would serve as a counterexample to the generalization about Irish people.

In the case of arguments, relevant counterexamples are denials of the conclusion that are consistent with the premises of the argument. Consider the following argument:

> If Sam studies all night, he passes the exam. Sam passes the exam. Therefore, Sam studies all night.

The conclusion of the argument is "Sam studies all night." The counterexample to the conclusion would be, quite simply "It is not the case that Sam studied all night". Is this a legitimate counterexample? If it is, the argument is invalid.

Let's think about the role of the counterexample in our reasoning about the argument: The first premise of the argument is that Sam would pass if he studied all night. Given that this is true, it does not mean that when we learn that Sam passed the exam we can legitimately conclude that he was studying all night. For all we know, he might have been playing video games all night but might have studied slowly and steadily in the weeks prior to the exam. Given that the premise

> If Sam studies all night, he will pass

is compatible with the counterexample to the conclusion,

> It is not the case that Sam studied all night

we are not forced to agree that he ***must*** have been studying all night. The existence of a counterexample to the conclusion that is compatible with the premises of the argument means that we should not accept the validity of the argument as a whole.

Sometimes, there are no counterexamples to the conclusion of an argument. If there are no counterexamples, if the conclusion always follows from the premises or the assumptions of the argument, then we say that the argument is valid.

7.4 Tautologies and Contradictions

There are certain special kinds of argument forms that are always true by virtue of their form and these are just as interesting and important. For example, it turns out that even in the most dire of circumstances, we can know with absolute certainty that the following statement is true:

(1) The plumbing system of the Kremlin is in Russia or it's not the case that the plumbing system of the Kremlin is in Russia.

Admittedly that's not a very exciting or informative claim. In fact, sentences of this form are always true. **Sentences that are true by virtue of their form are called tautological**. To show why, consider how we could replace the parts of the sentence that say something about the objects with variables. As we saw above, a variable is a letter or other symbol which can stand for some range of possible values. If I replace "The plumbing system of the Kremlin is in Russia" with the sentence variable *a*, I get the following:

(2) *a* or it's not the case that *a*.

Moreover, notice that no matter what sentence I insert for *a*, the total sentence that results will always be true.

(3) *a* or it's not the case that *a*.

(4) *Martians stole my mittens* or it's not the case that *Martians stole my mittens*.

Sentences of this kind are always true as a whole by virtue of their formal properties alone. Much of the study of logic concerns the structural or formal characteristics of reasoning. The process of thinking through the implications of what we believe or trying to decide whether we ought to believe one thing or another is a large part of the business of logic. The formal features of our assumptions often allow us to make claims based on other claims. Indeed, we can sometimes make those claims with absolute certainty independently of our knowledge of the world. My knowledge of astronomy is irrelevant to the truth of the following tautology:

(5) An asteroid cannot spin and not spin at the same time.

Tautologies are true no matter what the facts are. This means, they are always and necessarily true. It might be nice to know that we have access to such eternal truths, but tautologies are really not that useful. Tautologies are completely uncontroversial and uninformative and therefore are of virtually no interest to anyone except perhaps some philosophers. If I ask you what the weather will be like tomorrow and you tell me that it will either rain or it won't rain, your eternally true statement will not have given me any useful information. In any event, even if I am a philosopher, knowing (1–5) won't change any of my decisions or guide my actions in any significant way.

Practical or informal reasoning and argumentation rarely involves simple tautologies of this kind. Nor is it ever so certain. Ordinary conversations and arguments usually concern matters that are, or at least that appear to be, uncertain. Nevertheless, even in ordinary contexts, if you accept the premises of a valid argument, then you ought to accept the conclusion of the argument. Arguments fail to be *valid* when one or more steps in the chain of reasoning from premises to the putative conclusion do not obey the rules of logic. Invalid arguments are arguments where the conclusion of the argument doesn't necessarily follow from the premises. In other words, there are sometimes counterexamples.

Tautologies are sentences that are always true by virtue of their form, while contradictions are sentences that are always false by virtue of their form.

A sentence like

Jane graduated from Ohio State and she did not graduate from Ohio State.

is never true. The formal heart of a contradictory statement is easy to read off from an example like this. It is simply

A and $\neg A$

where "A" is a variable which can stand for any statement and "$\neg$" is the negation symbol. Putting the negation symbol in front of a statement is understood here to mean the denial of that statement.

7.5 Entailment

Philosophers and logicians often mention the idea of entailment when they introduce validity. Entailment is a relationship between a sentence and a set of sentences. In the case of a valid argument, we say that the conclusion is **entailed by** the premises. What logicians mean by entailment is subtle, but easy to explain. In order to understand whether premises entail a conclusion, we first think of the premises as a set of sentences and the conclusion as another sentence.

How do we determine whether one set of sentences (let's call that set Σ) entails another sentence? First, imagine joining the members of the set Σ with "and" to form a single larger sentence. **When two sentences are joined together with "and" the resulting sentence is called a conjunction**. For example, if Σ is the set of sentences:

> {Lubbock is in West Texas, Pigeons love cities, Red is a color}

and if when we join all the members of Σ with "and" we get the following conjunction:

> Lubbock is in West Texas and pigeons love cities and red is a color.

We can ask whether this conjunction entails some other sentence Γ, say for example, the sentence:

> Horses are larger than dogs.

The test we use to determine whether Σ entails that Γ is to consider whether the denial of Γ, in this case, the sentence:

> It's not the case that horses are larger than dogs.

and the conjunction of the members of Σ:

> Lubbock is in West Texas and pigeons love cities and red is a color.

when joined together with an "and" generates a logically contradictory statement. In this case, we are asking whether the sentence

> Lubbock is in West Texas and pigeons love cities and red is a color and it's not the case that horses are larger than dogs.

is a contradiction. It clearly is not. Given that the conjunction of Σ and the negation of Γ is not a contradiction, we can say that the truth of the members of Σ *does not* entail the truth of Γ. While the entailment relation does not hold in this case, it would hold for:

> "Pigeons love cities or Paris is in Ireland"

To see why, we can check to see whether the negation and the conjunction of the members of Σ lead to a contradiction. The sentence

> It's not the case that pigeons love cities or that Paris is in Ireland.

is logically equivalent to

> It's not the case that pigeons love cities and it's not the case that Paris is in Ireland.

or more intuitively,

> Pigeons don't love cities and Paris is not in Ireland.

Recall that conjoining the members of Σ gives us:

> Lubbock is in West Texas and pigeons love cities and red is a color.

Conjoining "Pigeons don't love cities and Paris is not in Ireland." with "Lubbock is in West Texas and pigeons love cities and red is a color." gives us the following contradiction:

> Pigeons don't love cities and pigeons love cities and Paris is not in Ireland and Lubbock is in West Texas and red is a color.

This sentence contains the contradictory assertion that pigeons love cities and pigeons don't love cities. Thus our original sentence "pigeons love cities or Paris is in Ireland" is entailed by Σ.

Exercises for Chapter 7

1. You can construct tautologies for yourself using "or," "not," and any assertions. What other logical words can you use to build tautologies and how would you do it?
2. What is a counterexample? Think of how we ordinarily use counterexamples to show that some argument fails.
3. What is the effect of contradiction in a conversation? If my conversation partner contradicts themselves in an obvious way, what usually happens?
4. In a valid argument, we say, if the premises are true, then the conclusion *must* be true. How should we understand the *must* in this claim? In what sense is it necessarily the case that the conclusion will be true?
5. Some people deny that there are any truths. What is wrong with flatly denying that any sentence is true?

8 Formal and Informal Fallacies

In Chapter 7, we introduced two of the most important virtues related to reasoning and argument, namely soundness and validity. With these in hand, we can begin to more clearly understand the most common examples of ways that reasoning can fail. As we shall see in the next two chapters, arguments can fail for reasons that are relatively easy to detect. These chapters are devoted to failures due to errors in formal reasoning. These are, in some ways, the easy cases. If we have some education in logic or probability theory, then we can detect obvious mistakes with ease. Formal errors in reasoning are so common that getting to the point of being able to recognize and avoid them is an important advance. The tougher cases are mistakes that we sometimes call *informal fallacies.* These are errors in reasoning that are due to violations of good practices in argument or inquiry. Some of these errors are so frequent and so seductive that they prove very challenging for most of us to avoid. However, even here we will begin to detect repeating patterns of reasoning that we can learn to avoid.

In Chapters 8 and 9 we will begin studying the formal fallacies. Chapter 8 introduces fallacies due to misunderstandings of logic while Chapter 9 explores fallacies in statistical and probabilistic inference. Chapter 10 presents the informal fallacies.

8.1 Introducing Fallacies

Originally, the word "fallacy" was associated with trickery or deception. Some of this original sense of the word persists in its modern usage. We say that a fallacious argument is one that is liable to convince a reader or listener to believe something for bad reasons. Nowadays, the term "fallacy" is generally used as a kind of catch-all term for failed arguments. The next three chapters will help you understand and identify fallacies with the hope that you can avoid being manipulated by them and can avoid using them inadvertently in your own thinking and decision-making.

Many texts in critical thinking present lists of fallacies to be avoided; *thou shalt nots* for critical thinkers. This can be helpful to a certain extent. However, it is important to understand that arguments can fail in an infinite number of ways. Thus, most texts list only the most common forms of fallacy. Rather than providing a finite list of examples of faulty kinds of reasoning, the goal of the next three chapters is to help you develop a sense for when reasoning is going astray and an understanding of how to return to the path of excellent thinking and decision-making. My working assumption is that it is easier and more effective to develop good habits than to expect to remember a long (incomplete) list of *thou shalt nots*.

What does it mean for an argument or a line of reasoning to go wrong? The kind of errors that have traditionally concerned philosophers are those where we fail to follow practices conducive to successfully seeking the truth. In Chapter 7 we say the two principal virtues of arguments: soundness and validity. Unsound and invalid arguments are those we wish to avoid. More broadly most philosophers and scientists are truth seekers and so it is little surprise that they hope to avoid obstacles to effective inquiry. One of the problems with fallacious reasoning is that it impedes productive inquiry.

One easy way of evaluating whether an argument or a line of reasoning is conducive to the pursuit of truth is to ask yourself whether what you are reading or hearing helps you to see the matter at hand more clearly or whether it encourages an emotional rather than a rational response. Another aspect to watch for is when you feel yourself relying on intuitive reasoning or "following your gut." In cases where emotion and intuition are steering your ship, it is highly probable that you are failing to make your reasoning process clear and are susceptible to biased or confused decision-making.

Inquiry and decision-making are closely connected to judgments about likelihood. For example, as we encounter new evidence, it should cause us to change our previous views about the world in ways that can be informed by the mathematics of probability theory. In Chapter 6, we saw how confirmation bias disposes us to overemphasize evidence that confirms our cherished beliefs and incorrectly ignore evidence that should cause us to revise or abandon them. Probability theory (as we will see in more detail in Chapter 9) offers insight into how we ought to take evidence into account in revising our beliefs.

We reason about likelihoods all the time. Take a simple example: Sam notices that a house has been on sale for over a year. There are several possible explanations for this, Sam reasons, perhaps the seller has set the price higher than market value; perhaps the house has structural problems, an unusual layout, a strange smell, or some other factors that have deterred potential buyers. As Sam learns more about the house, he will be able to adjust his level of confidence in any of those competing explanations. If he reads an inspection report that says the roof and foundation are in good shape, he will think that it is more likely that there is some nonstructural reason discouraging potential buyers. He won't know with certainty why the house sat on the market for a year, but as he learns more, his confidence in the likelihood of each explanation shifts relative to the others. Understanding how evidence influences our judgments about likelihood is crucial to thinking about evidence and inquiry.

Until recently, philosophers have focused primarily on logic and have paid relatively little attention to the kinds of errors in reasoning and decision-making that are associated with statistics and probability. In Chapter 9 it will be necessary to introduce some basic probability theory in order to explain some common fallacies related to probabilistic judgments. An elegant mathematical equation known as Bayes' rule figures prominently here. Bayes' rule shows how to correctly update one's beliefs in light of new evidence. While Bayes' rule is a simple mathematical statement, it has often proven difficult for nonmathematicians to understand and apply. While most of us will not actually apply Bayes' rule effectively in everyday reasoning, understanding the rule is still very useful. In Chapter 9, we will see how Bayes' rule captures what it would mean to reason effectively concerning probabilities as we weigh new evidence. The formal aspect of Bayes' rule is just a way of representing a disciplined method for thinking through decisions involving probability. Learning about Bayes' rule is not the final antidote for making errors in probabilistic reasoning, but it is very helpful nonetheless.

8.2 Powerful Rhetoric, Lousy Argument: The Slippery Slope

It should be noted that failing to follow the norms of good argumentative practice and good statistical reasoning can sometimes be compatible with successful persuasion. We can convince people who are not critical thinkers to believe things using terrible arguments. Some kinds of fallacies figure prominently in political and commercial rhetoric precisely because of their persuasive power. We will begin with one of the most common of these as an example of how fallacies can derail our thinking before devoting the rest of the chapter to exploring some simple formal fallacies.

As we saw in Chapter 1, political consultants and advertisers skillfully deploy fallacies or false claims in situations where it would be far more difficult to convince their audience using rational arguments or true claims. It would be impossible to rationally persuade an adult that eating breakfast cereal would endow them with athletic prowess, or that buying a particular phone would make them attractive, or mark them as a member of the creative elite. For the most part, advertising works by manipulating is in nonrational ways; by using fallacious arguments.

Likewise, politicians learned long ago that appealing to reason is highly inefficient and as a result, contemporary political campaigns in the United States are filled with fallacious reasoning and appeals to fear and other emotions. One fallacious pattern of reasoning that is used very effectively in political discourse is the so-called slippery slope. Slippery slope arguments are nonsequiturs that generally take the following form:

> If you pursue some course of action under consideration, then it will trigger a cascade of bad consequences ultimately leading to disaster.

So, for example,

> If you let Max skip his chores, then before you know it, he'll wind up in jail.

> If same-sex couples are permitted to marry, pretty soon people will be marrying their pets.

This fallacy depends on encouraging an audience to believe that some course of action would set in motion an inevitable chain of events. The metaphorical features of the slippery slope fallacy are compelling. It is easy to visualize oneself in the perilous circumstance of standing on a slippery slope. It conjures the image of losing one's footing and sliding uncontrollably toward disaster. The proponent of the slippery slope hopes that the fears of his audience members will overwhelm their critical capacities and that they will simply accept that the course of action under consideration will inevitably lead to distant but terrible consequences in the future.

The problem with the proponent's argument is that it does not actually provide reasons to believe that these disastrous consequences will come about. The proponent simply *asserts* that they will. In spite of being such lousy arguments, slippery slope fallacies are remarkably effective ways of persuading people. While they are rhetorically effective, they are easily detected and defeated by critical thinkers.

Why is it such an effective piece of rhetoric? To understand, it is worth considering the strategy being employed by the proponent of the slippery slope in some detail. Basically, the recipe runs as follows:

1. Create a scenario with some emotionally powerful negative features.
2. Illegitimately connect that scenario to the topic under consideration.
3. Claim that this connection is causal or that the emotionally charged scenario is an inevitable consequence of the course of action under consideration.

4. Hope that your audience is so emotionally overwhelmed by the thought of the negative scenario that they passively accept it as a consequence of the course of action or view under consideration.

With a little critical distance, we can see that slippery sloppy arguments are transparently manipulative and unconvincing. After reading the next page or so you will be forever immune from their persuasive power. Consider the following:

> If employers are permitted to force their employees to take drug tests, then pretty soon we will be letting corporations tell us what we can eat, when we should go to bed at night, and whom we can marry.

Now that you are alert to the rhetorical strategy, this argument should strike you as transparently weak and confused. The proponent of this argument is creating an illegitimate association between drug tests and mandatory diets, bedtimes, and choice of sexual partners. Arguments of this kind are mind-numbingly common in popular political discourse. Of course, the example under consideration will fail to rationally persuade a critical thinker that drug testing is a bad course of action. To begin with, a critical thinker will recognize that those who advocate the right of employers to conduct drug testing are simply not arguing for the right of employers to set mandatory bedtimes or diets. There is no reason to believe, at least no reason is presented above, that permitting drug testing leads to permitting mandatory bedtimes. There is no more reason to believe this any more than one ought to believe that permitting drug testing will lead to mandatory organic gardening.

You will often hear foolish journalists and moronic pundits in the media invoking slippery slopes and thereby signaling their inability to recognize slippery slopes as fallacious. You should be alert to some of the obvious symptoms. First, note the confidence and satisfaction with which such people deploy slippery slope fallacies. Second, learn to recognize the "next thing you know" marker of slippery slopes. Almost inevitably, this phrase indicates sloppy and confused thinking:

> Legalize weed and *next thing you know* they'll be legalizing heroin.

How many times have you heard this kind of argument? Such arguments are effective, in part because they play on our tendency to be lazy as thinkers. Pausing to think through the argument quickly reveals the error here. Proponents of marijuana legalization are not arguing for the legalization of heroin. Nor is any reason provided as to why legalizing marijuana will lead to the legalization of anything else. Likewise, proponents of marriage equality for gay people are not arguing for polygamy, bestiality, or the chance to marry household appliances. Conservative traditionalists are not the only ones to employ slippery slope arguments. They are common on all sides of political discourse:

> If pro-lifers have their way, they'll just turn women into baby-making machines with no choice of their own.

A moment's reflection should remind us that being opposed to legalized abortion does not mean advocating forced impregnation of women. The scenario of being forced to give birth is terrifying, but a critical thinker can separate the horrifying thought of being forced to give birth from the issue of the legal status of abortion. There are good arguments for and against preventing pregnant women from aborting their fetuses; fair minded and rational people can have honest disagreements on this question. In any event, it is certainly not the case that we should assume that all opponents of abortion have oppressive intentions towards women.

Slippery slope arguments are not only deceptive, but they also tend to encourage lazy and complacent thinking that undermines the pursuit of truth. Take, for example, the following:

> Prayer in school, Creationism in the classroom, the Texas Taliban is at it again, before you know it they'll be bringing back witch trials.

Proponents of school prayer or the teaching of creationism in school are not arguing for theocracy. By comparing proponents to the Taliban and by suggesting the slippery slope, the opportunity to engage directly with the merits of school prayer or creationism has been missed. The most important problem with slippery slope arguments is that they distract us from the real topic of discussion, replacing it with an emotionally worrying scenario that has no direct relationship with the issue itself. There are very good arguments against widespread drug testing in the workplace or the teaching of creationism in public school but slippery slope arguments are not among them. Fallacious reasoning is a distraction from the important business of figuring out what is true and trying to make good decisions.

Slippery slope arguments are common, they are rhetorically effective, and they are generally toxic to principled inquiry and the pursuit of truth. It is a deeply ingrained habit of thought that critical thinkers must resist. The easiest way to avoid it is to learn to recognize the clichés that accompany it: "next thing you know . . . ," "pretty soon everyone will be . . . ," "if you do . . . then you might as well . . . ," etc. The slippery slope fallacy is an example of the kind of distracting cliché that regularly bombard us in contemporary life.

8.3 Kinds of Fallacy

In the previous section, we saw an example of how the slippery slope fallacy leads us astray and impedes the pursuit of truth. Slippery slope is an example of an informal fallacy. As we saw above, fallacies can be roughly divided into two major types: the *formal* and the *informal fallacies.* Formal fallacies are arguments that fail by virtue of their logical or mathematical form. The remainder of this chapter introduces formal fallacies that are due to basic mistakes in logic. These are the easiest kinds of fallacies to detect once we have a little education in logic. In Chapter 9, we will examine formal errors in reasoning that are due to misunderstandings of probability.

In cases where the argument goes wrong for purely formal reasons, we do not need to concern ourselves too much with the facts of the case. Instead, we simply need to recognize that the argument is an example of an illegitimate pattern of reasoning. If we recognize that an argument is formally fallacious, we can legitimately reject it. Strikingly, we can reject formally incorrect arguments even in cases where we do not fully understand the content of the argument. Thus, **if you encounter a formally incorrect argument in a technical or scientific article you are entitled to reject the argument even in cases where you are not an expert in the relevant discipline.**

The second type of fallacy, the informal fallacies require a little more subtlety and good judgment. Many arguments will be formally correct and still fallacious. As we saw above, we call them fallacies because they are impediments to excellence in inquiry, reasoning, or decision-making. Informal fallacies convince their audience to reason poorly by means of trickery, emotion, or distraction. As we shall see, we can learn to detect and avoid informal fallacies with a little practice.

Informal fallacies can also result from our cognitive limitations. We are all limited in unavoidable ways. For example, our limited ability to pay attention or to hold relevant details in memory and can prevent us from reaching conclusions in the best ways possible. Sadly, none of us has an unlimited capacity to remember or attend to the details of long arguments. As we saw in Chapter 6, we face a range of other cognitive and psychological limitations that allow for fallacies to sneak in. It takes a special level of vigilance to avoid fallacies that result from our cognitive biases. Chapter 10 will introduce the informal fallacies in detail and offer some strategies to avoid them.

8.4 Formal Fallacies that Are Due to Misunderstanding of Logical Operators

Let's begin with arguments that fail by virtue of their logical form. The formal logical fallacies are those that result from errors that are due to misuse of words and phrases, such as "and," "or," "not," " if . . . then," "if and only if," etc. As we shall see, these words can serve as logical connectives and often serve to mark the logical structure of arguments.

Formal errors that result from misuse or misunderstanding of logical connectives are closely connected to the notion of validity. Fallacies involving logical connectives are straightforwardly invalid arguments. Recall from Chapter 7 that invalid arguments are those where it is possible for the premises to be true and the conclusion to be false. Conversely in a valid argument it is impossible for the premises to be true and the conclusion to be false. In later chapters, we will learn techniques that allow us to identify whether the form of an argument is valid. If an argument is not valid, then of course it is invalid. And if an argument is invalid, we should reject it.

There are an infinite number of ways for the form of an argument to be invalid, but for our purposes it makes sense to address the three most common ways that real arguments fail by virtue of their form. Frequently, arguments we encounter in everyday life will include the fallacies of *affirming the consequent*, *denying the antecedent*, and *affirming a disjunct*. In each case, the error is due to a misunderstanding of the logical role of logically important terms, such as "and," "if . . . then . . . ," "or," "not," and the like.

In later chapters, we will study additional set of fallacies resulting from misunderstanding the formal role of words, such as "all" and "some." Now, we will begin with cases involving the logical operators.

8.4.1 The Fallacy of Affirming the Consequent

The following example of fallacious reasoning is similar enough to examples that we have encountered that readers are likely to immediately diagnose the problem:

1. If my skin brushes against poison ivy, then I will have a rash.
2. I have a rash
3. Therefore, I brushed against poison ivy

This is another example of the logical fallacy of affirming the consequent. In this case, as with the cases we have seen in previous chapters the person making the argument is misled by the strong association that they have between poison ivy and rashes. While it is true that if I accidentally touch poison ivy while out hiking in the woods, then I will probably develop a rash, the fact that I have a rash should not lead me to conclude automatically that I brushed against poison ivy. Consider, for example, the possibility that I'm allergic to trail mix, or that new laundry detergent, or that I'm coming down with some peculiar disease. Given the existence of these counterexamples, it is entirely consistent with the truth of the premises of this invalid argument lines that the conclusion be false.

Let's treat this case a little more formally. Remember our definition for valid and invalid arguments; an invalid argument is one where it is possible for the conclusion to be false and the premises to be true.

This is clearly invalid by that standard. Given lines 1 and 2, I cannot be absolutely certain that line 3.

Explaining the name of the fallacy, *affirming the consequent*, requires a little background about conditional statements. Conditionals have antecedents and consequents. In the case of the conditional "if *P* then *Q*," *P* is the antecedent and *Q* is the consequent. Thus, for example, our conditional

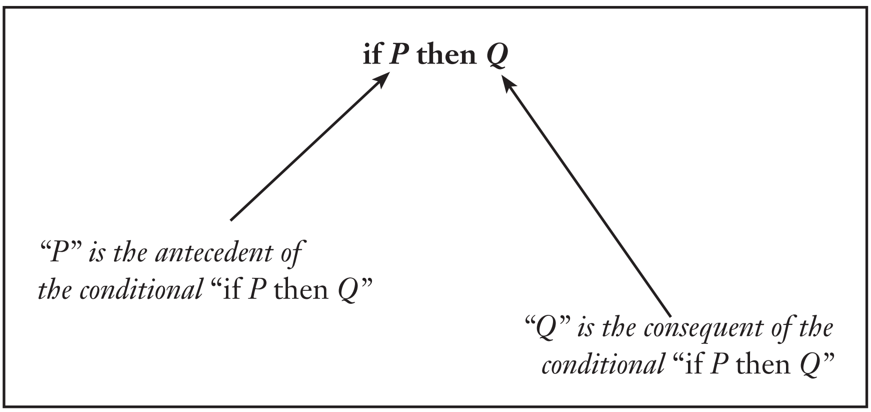
"if my skin brushes against poison ivy then I will develop a rash"

has as its antecedent:

my skin brushes against poison ivy (*P*)

and its consequent is

I have a rash (*Q*)



"Affirming the consequent" just means asserting that (in the case of our example) *Q*. The fallacy is called "affirming the consequent" because one begins from asserting the consequent and moves from there to affirming the antecedent. Notice that affirming the consequent by itself is not the problem with our fallacious piece of reasoning above. In fact, we assume that *Q* is true, it is one of our premises after all. The problem lies with the inference from lines 1 and 2 to 3. The inference is illicit insofar as it is a mistake to judge that the conditional and the consequent together license the conclusion that the antecedent of the conditional is the case. This mistake is due to a misunderstanding of the nature of the phrase “if . . . then.” Later we will see in detail how to understand the logical behavior of the “if . . . then” phrase. For now it should be clear that when we make a fallacious judgment like

if *P* then *Q*

and

Q

therefore

P

we are failing to acknowledge that the premises of the argument are fully consistent with the denial of *P*. Thus, if you’re only given

if *P* then *Q*

and

Q

there is no reason to conclude *P* rather than *not P*.

To return to our example, the fact that exposure to poison ivy is likely to cause rashes should not lead us to think that all rashes are caused by poison ivy. If you happen to have a rash it could have been caused by any number of things. So it’s consistent with having a rash and with the fact that exposure to poison ivy causes rashes that your rash is not caused by poison ivy. As we have seen, counterexamples would include allergies, diseases, and other possible causes.

8.4.2 The Fallacy of Denying the Antecedent

This fallacy also involves conditionals and is also symptomatic of a misunderstanding the role of the “if . . . then” phrase. Formally the fallacy looks like this:

if *P* then *Q*

and

not P

therefore *not Q*

to return to the case of the poison ivy, this fallacy comes out looking something like this:

1. If my skin brushes against poison ivy, then I will have a rash
2. My skin did not brush against poison ivy
3. Therefore, I do not have a rash

Remember that the antecedent of the conditional is "my skin brushes against poison ivy." The second premise is that argument involves the denial of the antecedent: "My skin did not brush against poison ivy" and as in the last case, the illicit move here is the inference from lines 1 and 2 to 3. As in the previous case the inference is unwarranted given that there are many ways that one can get a rash other than brushing against poison ivy. The counterexample to this fallacious arguments is a case where it's true that

if my skin brushes against poison ivy then I will have a rash

and it's true that

my skin did not brush against poison ivy

but it's false that

I do not have a rash.

As in the previous fallacy studied above, there are many other ways that I could have gotten a rash. I should not be so confident that I do not have a rash just because I was able to avoid the poison ivy.

8.4.3 The Fallacy of Affirming a Disjunct

There are a variety of words in English that can serve a logical function in the construction of arguments. Phrases in which the English word "or" serves a logical function are known as disjunctions in the parts of those phrases which are connected by "or" are called disjuncts. The word "or" poses one important challenge by virtue of a certain level of ambiguity. As we shall see in later chapters, there are two senses of the word "or," "exclusive or" and "inclusive or." Compare the following two sentences that feature the word "or":

a. In this store you can buy bread or cheese or eggs or any one of countless delicious products
b. Children, for desert you can have cake or ice cream

In the second case, we naturally assume that the speaker is offering the children a choice of cake or ice cream but not both. It might be clear from context or tone of voice whether this was in fact the intention of the speaker. If we are being very careful, we will indicate that the choice of cake excludes the choice of the ice cream and vice versa. We might do so by adding a few more words, as follows:

c. Children, for desert you can have cake or ice cream *but you can't have both cake and ice cream*.

By contrasting line c with line a we can see the difference between inclusive and exclusive senses of "or" quite clearly. It would be strange to encounter a store where you can only buy one thing. We would assume that the speaker meant an inclusive sense of "or." The inclusive sense of "or" simply means that when she says

a. In this store you can buy bread or cheese or eggs or any one of countless delicious products

the speaker does not intend to exclude the possibility that he could buy cheese *and* eggs (or any other combination of eggs, cheese, and delicious products) in the store.

Occasionally, the inclusive and exclusive sense of "or" are confused and this leads to fallacious reasoning. In English, you should assume that the intended sense of "or" is an inclusive "or" unless there's some clear contextual marker which leads you to think otherwise. So, for example, in English we sometimes try to indicate exclusive "or" through tone of voice, through the use of pauses in speech, or through the addition of explicit phrases as in line c above.

The fallacy of affirming a disjunct is a simple consequence of treating inclusive "or" as though it were an exclusive "or." An example of this fallacy runs as follows:

1. Sally holds an Irish passport or an American passport
2. Sally holds an Irish passport
3. Therefore, Sally does not hold an American passport.

The argument treats the "or" in the first premise is that were an exclusive "or"; as though it were actually claiming

> Sally holds an Irish passport or an American passport but not both.

It turns out, that in fact one can hold both an Irish and an American passport simultaneously since both the United States and Ireland permit dual citizenship between their countries. Therefore, the fact that Sally holds an Irish passport does not automatically exclude the possibility that she holds an American passport too. The argument above fallaciously concludes that Sally does not hold an American passport. Since the denial of this conclusion is consistent with the premises 1 and 2, the conclusion 3 does not follow validly.

8.4.4 Other Fallacies Due to Misunderstanding Logical Operators

The three cases we have considered so far are common examples of formal logical fallacies. As we have seen, they are invalid patterns of inference. Note, however, that all formally invalid arguments are fallacies and since there is an infinite variety of ways that the logical operators can be misunderstood, there is, therefore, an infinite variety of formal logical fallacies. Most of these have not been given names simply because they are not seen with sufficient frequency to require a label. The three cases that we have reviewed here appear very frequently and can be detected and avoided with relative ease.

The three cases we have covered here are common enough that you will be able to recognize them in a wide variety of contexts. One useful exercise is to begin reading newspaper editorials or listening carefully to political discussions on the radio or TV in order to find instances of these three fallacies. The reason that they are so common is probably due to the way that strong psychological associations can simply overpower our capacity to engage in deductive reasoning. The fallacies of affirming the consequent and denying the antecedent are clear cases where psychological association is pushing reasoning in a bad direction. However, given that these three cases are a small subset of the kinds of logical fallacies that we will encounter, we will need more a more general strategy for detecting invalid arguments. In later chapters, we will develop a set of techniques to determine whether an argument is valid or invalid. At this point, simply being familiar with the patterns exemplified by the three fallacies that we have introduced here is a good start.

Exercises for Chapter 8

1. Slippery slope arguments are fallacious, but they appear constantly in the media. Why do you think that they are so effective? Under what conditions could you imagine a slippery slope-style argument that is not fallacious?
2. Why do you think we tend to commit the fallacy of affirming the consequent and denying the antecedent so frequently?

9 Reasoning about Likelihood

You and I are fallible. We can be reasonably confident of many things, but in most cases it is possible that we are mistaken or that we were misled. Our confidence varies depending on how reliable the sources of our evidence are and how important the matter under consideration happens to be. In very important matters, we are more likely to seek greater levels of assurance than in less important matters. However, no matter how much we work to ensure that we are correct, we are almost always vulnerable to the possibility of error.

Confidence in our beliefs depends on how likely we think they are to be true. For example, I can be highly confident that I was born of a human mother, but I cannot exclude the very slight chance that I was not. I recognize that the likelihood that I was born of a human mother, given all the evidence available to me is very high. But I also recognize that it is possible that I could be wrong. Of course, if I am wrong about whether I was born of a human mother, I am wrong about a great many other things too. My whole picture of my life and my understanding of the social and natural world will have to be revised pretty fundamentally. Bizarre and improbable possibilities are fun to think about, but we should set very high evidential standards before we accept them.

Without some pretty extraordinary and convincing evidence, it would be foolish to take the possibility that I might not have been born of a human mother too seriously. While it is certainly possible that I am not a human, not all possibilities are equally likely. Part of being a rational critical thinker involves favoring those possibilities that are more probable over those that are less probable. Thus, critical thinkers need to know how to understand likelihood. This is the subject of this chapter.

9.1 Risks and Decisions

Many of our decisions involve calculating the likelihood of some outcome. From the trivial to the most consequential matters, there is some level of uncertainty in virtually all decision making. Given limited information we must decide the likelihood of different events or states of affairs. We weigh likelihoods in important matters, such as when we interpret the results of medical tests, when we make decisions concerning investments, when we evaluate risks to our safety, or when we make choices concerning our career path.

Even the most ordinary decisions usually involve some uncertainty and involve some reasoning about probabilities. If I am trying to decide whether I should lock my office door when I go to make a photocopy, my decision will depend on my beliefs concerning the likelihood of theft in my building at that particular time of the day. I realize that there is some risk that a burglar might steal my laptop during my brief absence from the office. When I make decisions, I must weigh the probability of theft against the minor inconvenience of locking and unlocking the door. Losing my laptop would be a severe blow to me, however, if the probability of a burglary is extremely small, I may decide that it is too much trouble to bother locking and unlocking the door and I will leave my door unlocked.

When we try to determine the likelihood of some event, we do so in the face of incomplete information. If we knew everything, and if we lived in a deterministic universe governed by a physics that looked like classical mechanics, we would never need to think about probabilities. However, we do not know everything and we live in a world that modern physics tells us is irreducibly random in some respects at the sub-atomic level. Thus, we are condemned both by our own ignorance and perhaps by the fundamental nature of physical reality to take probabilities into account when making decisions.

But what is a probability? **In contemporary mathematics and philosophy, we regard probability as the measure of likelihood**.

We talk about likelihood in situations in which we do not know for sure how things are going to turn out. When I role a die, I have no idea which of its six faces will face up. However, I do know that one of them almost certainly will. Barring an asteroid impact or the die being eaten by a pelican prior to its fall, or some other improbable event, the die will fall on the table in front of me. I am uncertain which side will face up. In spite of my ignorance about the precise outcome, I am right to think that I *might* roll a six. In fact, so long as the die is fair, I can be confident that I have about a 1/6 chance of rolling a six every time I roll a single die. I can be confident that I have a 50% chance that a fair die will land on one, two, or three, and not on four, five, or six.

Most future events that we care about involve some uncertainty. But this uncertainty comes in degrees. We do not know with certainty whether the sun will rise tomorrow, but it probably will. We do not know with certainty whether Harry will win the lottery tomorrow, but he probably won't. As a way of ordering more or less probable events, we can use numbers. The probability that an event will take place can be quantified on a scale between 0 and 1. If it is impossible that the event will take place, we say that it has probability 0. If it is certain that the event will take place we say it has probability 1. The higher the probability of some event is, the closer it will be to 1. The probability of any single player winning the lottery is very small (relatively close to zero). The probability of the sun rising tomorrow is very high (relatively close to one). The probability of a fair coin coming up heads, as we will see below, is half way between zero and one; 50/50 or 0.5. The probability of a single roll of a die coming up six is 1/6 or 0.16667.

Probability is the measure of likelihood:

The probability of an event can be quantified on a scale between 0 and 1.

If it is impossible that the event will take place, we say that it has probability 0.

If it is certain that the event will take place, we say it has probability 1.

The higher the probability of some event is, the closer it will be to 1.

In ordinary decision making, we are unavoidably confronted with challenges that require us to think about the probability of events. Take, for example, the following scenario: Let's say I am trying to decide whether to fly or drive to St. Louis for my cousin's wedding. Many of us are nervous about traveling by plane. We might worry about dying in a plane crash, but on reflection we might recognize that modern air travel is remarkably safe while driving in a car is significantly less safe. One's probability of dying in a plane crash is closer to zero than one's chance of dying in a car crash. At the same time, we might be concerned that our chances of surviving a plane crash might be significantly less than surviving a car crash. We might also weigh the risk involved in traveling against the cost of missing the wedding. Is going to the wedding worth the risk? Other price considerations might enter into our thinking. Even if traveling by plane is less dangerous than traveling by car, my decision might be sensitive to the financial costs. I might decide that the riskier, but cheaper option is preferable. In this way, I am putting a price on the risk that I am willing to take.

Clearly, deciding whether to drive or fly to one's destination is influenced by a variety of factors. Many of these factors will be subjective. How tolerant of risk am I, how important is it for me to please my cousin and their family, and so on. However, in addition to these subjective considerations, responsibly thinking through risks, requires that we become aware of objective features of reality. For example we might ask: How many people die in car accidents and plane crashes? How many miles are driven per year? Are accidents as common in all types of car or plane? These, and many other questions of fact, require some investigation and some might not have readily available answers.

How safe is air travel?

The International Air Transport Association gathers data on accidents involving commercial air travel. According to their results:

- More than 3 billion people flew safely on 36.4 million flights (29.5 million by jet, 6.9 million by turboprop).
- 81 accidents (all aircraft types, Eastern and Western built), up from 75 in 2012, but below the five-year average of 86 per year.

- 16 fatal accidents (all aircraft types) versus 15 in 2012 and the five-year average of 19 http://www.iata.org/pressroom/pr/Pages/2014-04-01-02.aspx.

Notice that these statistics do not include accidents involving private planes, military aircraft, and so on.

Some of these statistics are easy to find and are reasonably reliable, but others are not. For example, there is some consensus that almost 1.3 million people die in road crashes each year across the globe. There is less agreement as to how many miles are driven globally. More reliable statistics are available for the United States, but these numbers alone do not settle the issue. Even though we do not have reliable data for miles travelled versus traffic fatalities globally, even a rough comparison of miles flown per fatality to miles driven per fatality can lead one to think that flying is clearly safer.

Making sense of the facts about risks is not simply a matter of running a quick internet search. Critically thinking about statistics involves asking questions about the relevance of those numbers to the decision at hand. We might wonder, for example, whether the number of miles safely traveled is the right way to evaluate safety. The average journey on a commercial airliner covers many more miles than the average journey by car. Perhaps the right comparison is the number of journeys by both means. Furthermore, consider for instance that the most dangerous part of the flight is the take-off and landing. Given this, we might wonder whether we should think that a safe 1000 mile flight is twice as safe as a safe 500 mile flight.

Clearly, there are many questions to ask. Not only should we try to determine the best statistics, but we should also ask whether the statistics we are given are relevant to our central question: *Is flying really safer than driving?*

As we consider aspects of the decision we should also be aware of some of our biases. **Availability bias is likely to result in overemphasizing the risks involved in air travel**. Accidents involving commercial airliners are dramatic spectacles often involving large numbers of casualties. Because of their powerful emotional impact they will be repeatedly presented to audiences by ratings-driven media outlets in order to drive internet traffic and viewers to news sites and TV channels. Car accidents are less easily exploited for ratings. The fact that thousands of people die every day in car crashes is far more difficult to convert into a simple, spectacular narrative. Repeated exposure to a story about a single plane crash leads to our finding the idea of planes crashing easily available to us as we make our decisions. The result is that it is likely that the availability heuristic biases our judgments concerning the relative safety of air travel.

Even the most mundane decisions involve probability and statistics and many are subject to bias. When we decide between going to the park or going to the pool, our decision will involve beliefs about the probability of rain. When we think about risks, when we try to evaluate costs and benefits, and when we try to understand the best course of action to take under some set of circumstances, probability is almost always involved in our reasoning.

> Important personal decisions, for example, decisions concerning medical procedures, investments, and insurance are often, essentially, decisions concerning risk. Failure to correctly understand risk can have painful and expensive consequences.

As we shall see, in addition to decision making, judgments about probability are central to our ability to properly evaluate evidence. If, for example, I am trying to decide whether the scratching noises coming from my kitchen late at night are evidence of chipmunks, I must weigh the probability of chipmunks against the probability of mice. Strange scratching noises are compatible with chipmunks, mice, and demons. Which should I believe? Since there are more likely to be mice in kitchens than chipmunks or demons, in the absence of other evidence, it is more reasonable to assume that the noises come from mice.

In recent decades, psychologists and economists have demonstrated experimentally that people tend to make some very basic errors when they are reasoning about probability under certain circumstances. Once we come to recognize that we fail in predictable ways and that we fail for reasons that are deeply ingrained, we can take care to avoid, or at the very least mitigate these errors. The remainder of this chapter reviews the most prominent patterns of error concerning probability while offering some strategies for avoiding mistakes. The most effective strategies, as we shall see, involve understanding some of the basic formal features of probability theory.

9.2 Misunderstanding Statistical Independence

Casinos and government lotteries make substantial profits by exploiting our inability to reason clearly about probability. Casinos, for example, make most of their money by manipulating our difficulty thinking clearly about the phenomenon of **statistical independence**. This section introduces statistical independence, with the goal of inoculating you against faulty reasoning at the casino, in the stock market, and elsewhere. If this section succeeds, readers will save many times the cost of this book over the course of a lifetime.

In situations involving random events, we often make inferences about the future sequence of events that are unwarranted given basic principles of probability theory. Contexts in which randomness figures prominently include tossing a fair coin, games of chance such as roulette, blackjack and the like, and (to some extent) fluctuations in the value of equities in the stock market. The concept of randomness is philosophically and mathematically challenging. However, there are a few things we know with certainty about randomness. The first concerns the behavior of statistically independent events. The easiest example of statistically independent events is a series of coin tosses. We say that the coin tosses are independent insofar as the following fact obtains:

> **When I toss a coin (if it is a fair coin) the probability that it will come up heads is one in two or 0.5. This is the case independently of the history of the coin.**

The first point to notice is that each time I toss the coin, the odds that the coin will come up heads is 0.5; it is a simple fact that each toss of the coin is statistically independent from others. The fact that I toss a coin after tossing the coin previously does not change the probability that when I toss the coin the odds that the coin will come up heads is 0.5. This simple principle is extremely hard for us to accept in practice. **Imagine watching a fair coin land on heads 10 times in a row. It would be extremely difficult for most of us not to *feel* that the odds of the coin landing on tails in the 11th coin toss is greater than 50%**. However, this very strong feeling is misleading. Given that each coin toss is statistically independent, even in the unlikely event that the previous sequence of coin tosses had landed on heads 10 million times, the odds of landing on heads or tails remains 50%. In spite of our strong inclination to believe otherwise, future coin tosses are statistically independent of past coin tosses.

Most of us will decide whether or not to place a bet, say for example, at the roulette wheel or in the stock market based on our past success or failure. If I feel that I am on a winning streak, I may be inclined to bet more on the next round of play in order to capitalize on my luck. Alternatively, if I have been winning for a while, I might be inclined to be more conservative or cautious given my belief that my luck might be running out or my desire not to push my luck. Notice that this way of thinking about streaks and luck at the casino is based on a very simple misunderstanding of the independence of the probabilities involved in each play of, for example, the roulette wheel. Failure to recognize the probabilistic independence of each round of play accounts for these kinds of fallacious inferences.

One way that we can free ourselves of fallacious reasoning about statistically independent phenomena is to try to account for our feeling that past tosses of the coin influence the odds of future tosses of the coin. Returning to our example above imagine the following sequence of coin tosses:

. . . T, H, T, T, H, H, H, H, H, H, H, H, H, H

We can very easily find ourselves arguing something like the following:

1. 10 times in a row the coin toss came up heads.
2. This is highly unlikely, and this string of heads is highly unlikely to continue.
3. Therefore, it's more likely than not that the 11th coin toss will come up tails.

This piece of fallacious reasoning is known as the **Monte Carlo Fallacy**, or the **Gambler's Fallacy**. It is the mistaken inference from the past behavior of some statistically independent phenomenon to some claim about the probability of a future statistically independent phenomenon.

We can overcome the natural inclination to commit this fallacy by asking ourselves whether the coin, or the roulette wheel, knows anything about past coin tosses or spins. Clearly, the coin or the roulette wheel (if they are fair and not tampered with) has no memory of what has happened in the past. Our fallacious reasoning assumes that there is some kind of mechanism in operation connecting future coin tosses with a history of past coin tosses in some relevant way. If there were such a mechanism, then it would not be a fair coin. Given that there is no such mechanism, the best strategy is to treat each coin toss as statistically independent.

Once we understand statistical independence, we can more easily see why, for example, the fact that someone has had a run of good luck at the casino, or playing the lottery, has no influence on their chance of winning in the future.

9.2.1 Seeing patterns and the clustering fallacy

Human minds are eager to find patterns. Therefore, for example, lying on the grass and looking at the sky, it is hard to avoid seeing faces and shapes in the clouds. Pattern detection is a vital feature of our cognitive capacity, but when we start finding patterns everywhere in unhelpful ways; finding the face of a religious figure on a slice of toast or hearing hidden messages in the static on the radio, we call this tendency **pareidolia**.

Pareidolia distracts us by finding patterns that are causally irrelevant. The faces we find in the clouds or on our toast may be entertaining, but their appearance or existence should not guide important decisions.

How can we tell whether we are seeing patterns that matter? Some guidance from probability theory can help. Imagine tossing a coin some number of times and observing some sequence:

. . . , TTHHTHTTTHTHTHTHTHTHTHTHTHTHHHTHTHTTHHTHHH . . .

We notice that there is a striking pattern of alternating heads and tails in this sequence. What should we make of it? We remind ourselves that each coin toss is statistically independent, we understand that this pattern should not be taken to indicate anything significant. **Note that given a sufficiently high number of repeated coin tosses, we should expect to see patterns like this at some point**. Note also that while we can confidently expect patterns to appear, we are not able to tell *when* this pattern will appear.

9.2.1.1 Texas Sharpshooter Fallacy

Someone who draws a circle around this pattern and takes it to indicate something important is committing what we call the **clustering fallacy** or more colorfully, the **Texas sharpshooter fallacy**. This fallacy gets its name from the story of a Texan who claims to be an excellent marksman. In fact, he is not a marksman at all. Instead, he simply shoots at the side of a barn in a relatively random manner. After a certain number of shots, he locates a cluster of bullet holes, draws a circle around the cluster and invites his friends to congratulate him on his accuracy.

The fallacy involves cases in which some random cluster of features is illegitimately ascribed some kind of significance or causal relevance. The story of the Texan marksman involves deliberate deception. The most interesting cases of the clustering fallacy do not involve the conscious intention to deceive. In fact, most of us have a tendency to ascribe significance to some cluster that we encounter in data or to some streak we find in sequences. The trouble is, in many cases, we over-ascribe significance to such clusters.

If, for example, we were to find an unusually high incidence of some disease associated with a school or neighborhood, it is incumbent upon us to look at the entire set of evidence in order to make sure that the cluster in question is not solely an artifact of a random distribution. For example, in any random distribution of dots on a map, we are likely to find some cluster of dots.

Knowing that we should expect to find clusters in random distributions does not make it any easier to stop seeing those clusters as having potential significance. I once worked in a building where three former workers had died of brain cancer. It is difficult to avoid inferring some causal relationship between the conditions of the building and the cases of cancer. However, focusing on the cluster apart from the rest of the data is close to what the sharpshooter did when he drew a circle on the barn. A geographical clustering of cancer cases, without any other evidence, is likely to be the result of the random distribution of cancer cases. If brain cancer were randomly distributed throughout the population, we should expect many such clusters.

9.2.2 The law of large numbers

The law of large numbers is a mathematical theorem that tells us that we can know some important facts about large numbers of random events. In the case of statistically independent random events, like coin tosses, rolls of a die, or spins of the roulette wheel, given a sufficiently large sequence of events, we can predict some features of how things will go. Specifically, we can know that the average value of the outcomes gets ever closer to the **expected value**. The expected value is the value we would expect by reasoning a priori. Therefore, for example, in the case of coin tosses the number of heads and tails come closer to being equal the larger the sequence of coin tosses. We would expect that the number of heads would be almost the same as the number of tails given enough tosses of the coin. Likewise, the given a fair six-sided die, we could reason a priori that the chance of it coming up three, for example, is one in six. Given a sufficient number of rolls, we will find threes appearing around one sixth of the time. If we were to roll the die 600 times, its value would be three around 100 times, not exactly 100 times most of the time, but close.

Notice what is being asserted here. From the comfort of our armchairs, without actually rolling the die, we can reason a priori that the average roll of a fair six-sided die will have the value 3.5 given a sufficiently large number of repetitions—**the expected value for a roll of the six-sided die is 3.5**. The law of large numbers is a mathematical result (also the kind of result we can derive from the comfort of our armchairs) that assures us that the average value of the rolls will converge to (1+2+3+4+5+6) ÷ 6 or 3.5. If we were to run an experiment to see whether, in fact the die behaves as we expect over time, we will be vindicated. You are welcome to try this for yourself with coin tosses or rolls of the dice, but there is no need, since the law of large numbers guarantees that this will be the case. It is a remarkable fact that we can know, via our understanding of the law of large numbers theorem, that this will be the case. We will not review the proofs of the theorem here (you can easily find a variety of proofs online) but instead, we will examine some of its implications.

The law of large numbers describes a feature of random events that are statistically independent. Given a sufficiently large sample, the anomalous streaks of good luck or bad luck that excite gamblers at the casino will be smoothed out over time. What this means is that the casino may lose money on some occasions, but in the long run the law of large numbers guarantees that they will profit in a predictable manner. However, as we saw previously, any one roll of the die or spin of the roulette wheel will be statistically independent from all the others.

> When we reason about risks and randomness, **it is important to remember that it is highly likely that unusual things will happen**. You can count on odd-looking patterns, coincidences, strange streaks, and other kinds of outliers in a sufficiently large sample of random phenomena. You can be certain to find some people doing unusually well in random games. The fact that they did well in a random game is not an indication of their virtue or of their having a better strategy than others.

9.2.3 The regression fallacy

Imagine that on average there are 12 burglaries in a neighborhood per year. One month there is an unusually high number reported, say 5. The neighbors decide to take action. They install signs in the neighborhood alerting potential burglars that they have an active neighborhood-watch program in effect. The following month there are no burglaries. The neighbors are delighted and conclude that the signs served as an effective deterrent to crime and recommend that they be installed in other neighborhoods. Without any additional information, it is likely that the neighbors are committing what is called the **regression fallacy**.

The regression fallacy is the failure to recognize that things normally go back to normal after anomalous or unusual events. Without any additional information, we should assume that the rate of burglaries will return to normal levels independently of posting the sign. We can be confident of this in virtue of the fact that statistically independent events tend to regress to the mean. Therefore, ascribing causal efficacy to the sign is fallacious.

Let's think about what causes us to commit the regression fallacy. To begin with, it is generally the case that we take note of anomalously high levels of some event. If there is a spike in the number of car accidents, burglaries, students failing a course, or some other undesirable outcome, it tends to attract our attention. If these anomalous events are undesirable, it is difficult to resist the urge to take action to try to reduce them. However, in events that are partially the result of randomness, our sense that "something must be done" is not necessarily grounded in a correct understanding of the facts. As we saw above, in any sequence of events that are at least partly the result of randomness, we can be confident that there will often be occasions when we see spikes or natural fluctuations in the rate of those events. These spikes happen, but we tend to see the value of the events returning to their normal rate soon after. The mistake that the neighbors made in our example was to assume that the signs were causing the reduced rate of burglaries rather than seeing the trend as the result of a predictable statistical phenomenon. In statistics, it has long been recognized that measurements of some value that give some extreme result tend to be followed by measurements that are more normal or closer to the average.

To speak very loosely, the regression fallacy involves thinking that our actions are causing things to return to normal instead of realizing that *most of the time things tend to be normal*. Avoiding the regression fallacy is very difficult. However, it is wise to resist the urge to try to fix things before determining whether the problem we are seeing is the result of normal statistical variation. In the case of events that are partially determined by randomness, the most efficient strategy to adopt in the face of some undesirable spike might be to simply do nothing.

9.2.4 Conjunction fallacy

The work of Amos Tversky and Daniel Kahneman has appeared prominently in this book. In Chapter 6, we saw how they were among the first to point out the regular patterns that we find in errors in reasoning about probability. One of the most famous examples of such an error is the so-called Linda cases or the conjunction fallacy. When we commit the conjunction fallacy, we are incorrectly judging the likelihood of a conjunction of two events as greater than either of those events in isolation. In some cases, we are disposed to judge the probability of two events A and B as being greater than the probability of A by itself. There is something very odd about this mistake. Clearly we cannot, on reflection, accept that A is less likely than A and B, but in the case of the conjunction fallacy, this is exactly what we judge to be the case.

In their experiment introducing the conjunction fallacy, Tversky and Kahneman present their subjects with the following short description of Linda followed by a task in which they are asked to rank the likelihood that she participates in some set of professions, hobbies, or other activities. What follows is the description of Linda, the instructions that they gave to subjects and the mean ranking that their subjects gave to the statements describing Linda:

> *Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.*
>
> *Please rank the following statements by their probability, using 1 for the most plausible and 8 for the least plausible*
> *(5.2) Linda is a teacher in an elementary school.*
> *(3.3) Linda works in a bookstore and takes yoga classes.*
> *(2.1) Linda is active in the feminist movement. (F)*
> *(3.1) Linda is a psychiatric social worker.*
> *(5.4) Linda is a member of the League of Women voters.*
> *(6.2) Linda is a bank teller. (T)*
> *(6.4) Linda is an insurance salesperson.*
> *(4.1) Linda is a bank teller and is active in the feminist movement. (T & F)*

Notice that subjects ranked "*Linda is a bank teller and is active in the feminist movement*" as being more probable than "*Linda is a bank teller*". Now that we have introduced some probability theory, the problem with this judgment is obvious. However, the appeal of "*Linda is a bank teller and is active in the feminist movement*" is also clear. Tversky and Kahneman diagnose our tendency to commit this fallacy as being due to the mistaken use of the representativeness heuristic. While there are a variety of interpretations and explanations of the Linda cases, it is clear, no matter what the explanation, that we would not commit the conjunction fallacy if we were sensitive to the simple fact that for any two events A and B, A is more likely than the conjunction of both events A and B.

9.2.5 The prosecutor's fallacy

The prosecutor's fallacy is a mistake in reasoning where probabilities are incorrectly conflated. Consider the following example: A red-bearded man with a Jeep Wrangler was spotted at the scene immediately before some crime was committed. Since Sam is a man with a red beard and a Jeep Wrangler, the police take him in for questioning and the prosecutor is delighted. His statisticians inform him that the chance of any one person having both a Jeep Wrangler and a red beard is staggeringly small. It is highly unlikely that any person chosen at random would fit this description.

However, the prosecutor's fallacious inference lies in his inference from the staggeringly small odds that Sam has both a Jeep Wrangler and a red beard to the conclusion that Sam is the perpetrator.

To see why he has made a mistake, we first need to think about how many people fit the description. This will vary depending on the sample of people under consideration. Let's say for instance that the crime takes place in a large metropolitan region like Los Angeles with a population of around 9.8 million people. How many people have Jeep Wranglers and red beards in LA? We can begin to make a rough estimate given that we know there are 5.8 million vehicles registered in LA County, in the United States where 2–6% of the population has red hair. These are not precise numbers, but they are the kind of statistics one can find from reasonably reliable sources on the internet or at your local public library.

Given these numbers, we can estimate that there are therefore between 116,000 and 348,000 cars in LA County owned by red-haired people. Let's assume that half of these cars are owned by red-haired men; 58,000–174,000. In the United States, slightly more women than men hold drivers licenses at present, however the numbers are close enough to 50% as to make little significant difference in our conclusion.

It is a little difficult to estimate the number of men with beards, and there is a great deal of variation in this percentage over time due to changing fashions. One study noted that in 1970, 58% of British men had beards. It is unlikely that the percentage is anywhere near this high in LA county at present.[1] Let's assume that only 10% of men are bearded. This means that there are between 5,800 and 17,400 red-bearded car owners in LA county.

How many of these red-bearded car owners drive a Jeep Wrangler? This is a difficult question, but we can take a stab at it by thinking about the number of Jeeps on the road as a percentage of the total number of vehicles. Let's say that 194,142 Wranglers were sold in the United States in 2012 (various Internet sources suggest that this is the correct answer). Roughly 9 million new passenger cars were sold in the United States that year, so this means that roughly 2% of cars sold that year were Jeep Wranglers. Let's assume that this percentage is roughly stable, although again it undoubtedly fluctuates with fashion. Again, we do not need precise numbers for our purpose here.

Given the estimate of the number of red-bearded car owners in LA County (between 5,800 and 17,400) and calculate the number of these men who own Wranglers (2%). We now have a total population of between 116 and 348 people.

While this is undoubtedly a flawed, back of the envelope calculation, it should suffice to make a reasonable prosecutor less confident that Sam is a legitimate suspect. Given our calculations, Sam should be considered as having at most a 1in 100 chance of being the suspect.

[1] Robinson, Dwight E. "Fashions in shaving and trimming of the beard: The men of the illustrated London news, 1842–1972." *American Journal of Sociology* (1976): 1133–1141.

Thus, taken by itself, the fact that he fits the description is insufficient evidence to convict him.

The prosecutor's mistake is to miss the fact that any one of those 116–348 people could have been the guilty party. Thus:

> ***While it is true that having a Jeep Wrangler and a red beard is highly unusual, there are many people in the city with this highly unusual set of features.***

The prosecutor's thinking is clouded by the powerful impact of the first statistic. The prosecutor is so impressed at how unlikely it is that someone would have both a Jeep Wrangler and a red beard that he fails to notice that this simply puts Sam in a reasonably large pool of suspects. Given this pool of suspects, the likelihood of Sam's being guilty is not so overwhelming or impressive.

9.3 Misunderstanding Conditional Probability

As we have seen, cognitive bias plays a detrimental role in some important decision-making contexts. Misapplication of the representativeness and availability heuristics can lead us to make straightforwardly bad judgments. One of the most challenging areas for us concerns judgments involving conditional probability. This section introduces unconditional and conditional probability formally before turning to some important ways in which biases lead us astray.

9.3.1 Unconditional probability

When we think about the probability of events, it is helpful to put numbers or odds on their happening. As we saw above, it is common to think of probability as a quantity between 0 and 1:

- If there is no way that something will happen, we can say that it has a probability of 0.
- If something will happen with absolute certainty, we can say that it has a probability of 1.
- We say that uncertain events have a probability of between 0 and 1.

Therefore, in the case of the fair coin toss in the previous section, the probability that the coin will come up heads is 0.5.

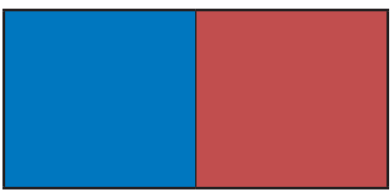
Let's say that we have tossed a coin in the air. Let's exclude the possibility that it will be destroyed, stolen by aliens, or otherwise interfered with as it falls back to earth. We are assuming that there are two events; heads (A) and tails (B) both of which have a probability of 0.5

$P(\mathrm{A}) = 0.5$
$P(\mathrm{B}) = 0.5$

Coin tosses are **mutually exclusive** events, meaning that if the coin lands on heads, it cannot land on tails. When we say that a tossed coin has to land on either heads or tails we are asserting:

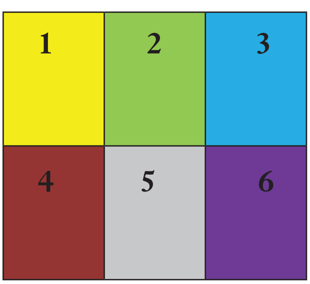
$P(\mathrm{A}) + P(\mathrm{B}) = 1$

The idea is that the two options; *heads* and *tails* exhaust the space of possible events for coin tosses. This idea of an exhaustive space of possible events for mutually exclusive events is known as a **sample space**. In geometrical terms, we can imagine the set of possible futures for the coin flip being as follows:

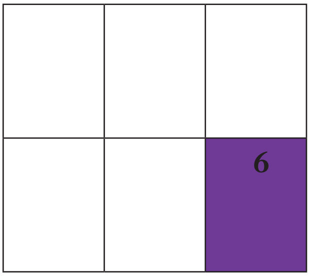


Heads Tails

The sample space for rolls of a six-sided die looks like this:



Once again, there is an equal chance that any of these six futures will obtain. If we were interested in how likely it is that we would roll a six, we can represent the chance geometrically as a portion of the sample space as follows:



The idea of treating probabilities in terms of sample spaces is simply a way of representing the set of all the possible ways that things could be, and dividing them up according to the mutually exclusive alternatives. For example, if I am concerned with being struck by lightning over the coming year, I could consider all the possible ways that things could be. There is some probability that you or I will be struck by lightning. According to the National Weather Service of the United States the odds that a person in the United States will be struck by lightning during any given year is around 1/775,000 or $1.29032258 \times 10^{-6}$ or .00000129032258. This is called an unconditional probability.

$P(\text{L}) = 0.00000129032258$

If we were to attempt to map out the portion of the sample space containing the set of futures where I am and am not struck by lightning, the portion where I am struck by lightning would be very tiny indeed. On a computer screen, it could be represented as a single pixel in a box roughly 880×880 pixels in size.

What we mean by an unconditional probability is simply calculating the likelihood of an event based solely on the number of events under consideration divided by the number of events. This is sometimes called the **relative frequency** or **a posteriori** approach to probability. In this case, the National Weather Service simply divides the number of people in the United States by the estimated number of deaths and injuries caused by lightning strikes in a year.

Notice that in the case of the coin toss, we did not determine the value of the probability that the coin would come up heads or tails by studying a large sample of events. Doing so, in practice, would probably not have resulted in a precise value of 0.5 for *heads*; if we toss a coin a large number of times, the number of heads and tails are highly unlikely to be the same. However, the larger the number of coin tosses, the closer we approach 0.5 for heads.

We arrived at the 0.5 value for *heads* **a priori**, by assuming heads and tails to be equally likely outcomes from a pair of types of event that exhausts the space of possible outcomes.

In the case of a six-sided die, we can reason our way to the probability of the die coming up 6 given that we know that each of the six sides has an equal chance. We will use the label D6 for an event in which the die landed with the 6 side facing up.

$$P(\mathrm{D6}) = \frac{\text{The number of ways D6 can occur}}{\text{The total number of possible outcomes}}$$

Therefore, a priori, as we saw above, we would say that $P(\mathrm{D6}) = 0.1666$ . . . or 1/6.

9.3.2 The probability of two events happening together: multiplying probabilities

Suppose we want to know the likelihood of two events *A* and *B* happening together. Say *A* is the probability of a fair coin coming up heads, and *B* is the probability of a die coming up six.

We know that $P(A) = 0.5$ and, as we saw above, we know that $P(B) = 0.1666$.

Two determine the likelihood of both events happening together, we multiply the value of their probabilities. In this case:

$.5 \times 0.1666 = 0.0833$

(One-half multiplied by one-sixth is one-twelfth.)

The chance that I would both roll a six on a die and that a fair coin would come up heads is one in 12.

9.3.3 What is conditional probability?

The National Weather Service is a reliable source of information. As we saw above, they tell us that if you live in the United States, the probability of your being struck by lightning this year is 1/775,000. This is an interesting piece of data, but by itself it is not particularly informative. It is unlikely that such a piece of information would help you to make better decisions. Let's say that you learned that people are more likely to be struck by lightning when they are not driving to work. People who drive to work are probably safer from lightning strikes than people who are not. You might have learned that being inside a car provides excellent protection against lightning strikes since the body of a car acts as a Faraday cage, sending electrical currents safely to the ground while protecting the interior of the car.

If I were concerned about being struck by lightning, I might want to know exactly what the conditional probability of being struck by lightning is given that I drive to work. People who drive to work have a lower probability of being struck by lightning than the unconditional probability of being struck by lightning. While we might guess that people who walk to work have a higher probability of being struck by lightning than either the unconditional probability or the conditional probability for drivers.

To take another example, we all have some unconditional probability of tearing the anterior cruciate ligament (ACL) in our knees. However, one might be interested in knowing the conditional probability of tearing the ACL given that one plays soccer. In order to answer this question, it would be necessary to restrict one's attention to the set of soccer players. Then, one would simply need to determine the likelihood of a torn ACL in this subgroup of the population. This would provide the conditional probability of a torn ACL given that one plays soccer. The conditional probability of tearing one's ACL given that one plays soccer is far higher than the unconditioned probability. One might discover that the conditional probability of tearing the ACL is ten times as likely (I have no idea what the actual statistics are) given that one plays soccer than the unconditional probability.

Knowing the conditional probability of being struck by lightning under different circumstances can be a useful guide to action, certainly to choice of hobby. The way that conditional probability is usually presented formally can be explained by example. The probability of event A given event B is usually written as follows:

$P(A \mid B)$

Conditional probability is definable in terms of unconditional probability as follows:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Most of our most prominent and costly errors in reasoning about probabilities come from confusions related to conditional probability. For example, it is very rare that $P(A \mid B)$ is equal to $P(B \mid A)$. For example, given that you were hit by lightning, there is a high probability that it was a cloudy day, but the probability of being struck by lightning on a cloudy day is very small.

Let A be the cloudy day and let B be getting struck by lightning:

$P(A \mid B) = \text{high}$

$P(B \mid A) = \text{low}$

While this might seem obvious, confusions arise in cases where emotions run high. For example, most medical tests will have some rate of so-called false positives. What this means is that the test will sometimes report that a patient has some disease when, in fact they do not. A test for some very rare disease D might have the following characteristics:

In 90% of people with the disease, the test is positive.

In 5% of people without the disease, the test is positive.

If 20 million people are tested, 1 million will receive a positive test result. Getting a positive result is a disturbing outcome, but should it cause you to believe that you are sick? Strikingly, if a healthy person were screened every year for 60 years, then the chances of receiving a false positive are approximately 95%. Therefore, given frequent testing, one should expect a false positive.

This is not merely a hypothetical scenario: One of the most important medical tests currently available, the Papanicolau (or pap) smear, has a similar profile. The pap smear detects changes in cervical cells related to cancer. It is one of the most effective cancer screening tools available. According to some studies, it has reduced cervical cancer mortality by as much as 99% in women who are tested (DeMay, 2000). If one has, or is in imminent danger of developing, life-threatening cervical cancer then the test is highly likely to give a positive result. There is a small rate of false negatives, but notice that this is a risk only for the very small number of people with the disease (roughly 2% of women). The risk of false positives, by contrast affects almost all women (the healthy 98% of women). Healthy women are estimated to have between 1% and 10% chance of receiving a false positive result.

This might sound like a relatively small probability of a false positive, but given that all women are advised to regularly undergo a screening throughout their lives, the implications of this rate of false positives are quite striking. One researcher offers the following analysis:[2]

> Assuming a 5% false-positive rate and screening 100 million women annually, then 5 million false positive test results would be expected, incorrectly identifying 5 million healthy women as having squamous intraepithelial lesions, when the actual number of women with such lesions is thought to be closer to 2 million. At the individual level, if a healthy woman were screened every 3 years from age 18 through her mid-70s, for a total of 20 Pap smears, and the false-positive rate were 5%, there is an almost 2 in 3 chance (64.2%) that she would experience at least 1 false positive result. (If screened every year, for a total of 60 Pap smears, there is approximately a 95% chance of at least 1 false-positive result.) Thus, the risk of a woman's being incorrectly identified as having a squamous intraepithelial lesion may be substantially greater than the risk of actually having a lesion and far and away exceeds her risk of dying of cervical cancer (DeMay 2000).

Careful consideration of the rates of false positives should lead us to conclude that healthy women should not be surprised or unduly alarmed by receiving a positive result on their screening. In fact, all

[2] DeMay, Richard M. "Should we abandon pap smear testing?" *American Journal of Clinical Pathology. Pathology Patterns Reviews*.114. Suppl 1 (2000): S48–S51.

women should expect a positive result at some point in the course of their lives. A positive result should lead one to inquire further, but it should be understood in light of the probability of false positives.

9.4 The Importance of Base Rates

Errors of the kind discussed above are instances of what is known as the **base rate fallacy**. When we commit the base rate fallacy we incorrectly neglect an important part of the background information in some context of decision making. Judgments about probability are always judgments of relative likelihood. Therefore, when we hear that a test detects the presence of a disease in 99.9% of those tested who carry the disease, the impressive number alone does not allow us to precise interpret the significance of a positive result. Such judgments require attention to background information.

To understand the role of background information, consider the following assertion:

> 30 of the students who passed the exam read the study guide for the course.

Initially, you might be tempted to think that this is evidence that reading the study guide is helpful. However, this number by itself is of relatively little significance. In order to understand the meaning of the statistic, we need to know additional background information. For example, if you knew that

> 400 students took the course and 270 passed the exam without reading the study guide

you might be less impressed by the usefulness of the guide. Furthermore, imagine if you learned that

> 60 students who failed the exam had also read the study guide.

Your attitude toward the original claim about the 30 students who passed the course would be quite different. In fact you would be entitled to believe that reading the study guide could be harmful to your chances of success.

The point here is that the context in which statistical evidence is presented determines the meaning of that evidence. The phrase "base rate" is intended to capture part of this context. Strictly speaking, the base rate is the rate of some value of interest prior to the new evidence, information, or intervention under consideration. In order to know whether to use the study guide for the course, one must first understand how students ordinarily fare independently of using the study guide. From there, we can judge the positive contribution that results from adding the study guide for the course. In other words, paying attention to how things normally go is essential for understanding whether some intervention is effective.

9.4.1 Informal reasoning, conditional probability, and base rates

Conditional probability generates real challenges for reasoning and decision making in circumstances where base rates are neglected. Consider the following scenario: A patient has torn the anterior cruciate ligament (ACL) in his knee. In the past, a torn ACL was likely to be a career-ending injury for American football players and, in fact, torn ACL's are far more likely to affect football players than nonfootball players. The question for us is how likely is it that the patient we are encountering is a football player? In this example, three different values are relevant to answering the question:

The percentage of football players in the general population

The unconditional probability of a torn ACL for members of the general population

The conditional probability of a torn ACL given that one plays football

In the simple example just given, I did not provide numerical values for the probabilities under consideration. Once the numbers are assigned for each of these three values, there is a straightforward way to calculate the likelihood that the patient is a football player. We will discuss how one does this in detail below. However, without some exposure to basic probability theory we will consistently fail to evaluate our set of options correctly. Specifically, most of us will overestimate the likelihood that the torn ACL belongs to a football player. Consider that active football players comprise a very small percentage of the population. Consider also that a small (but nontrivial) percentage of football players will suffer a torn ACL. There are roughly 80,000 cases of torn ACL every year in the United States, women are more susceptible to the injury than men, and surprisingly, 70% of tears occur in noncontact events.[3] In spite of this, the association between football and torn ACL is strong enough to lead most of us to overestimate the likelihood that the patient is a football player.

9.4.2 Base rates and the representativeness heuristic

Even in cases where we have some acquaintance with probability and are given explicit numbers, we are susceptible to error. In the early 1970s, Amos Tversky and Daniel Kahneman discovered that our judgments in cases like these are systematically flawed. For example in the ACL case, an American doctor is very likely to overestimate the likelihood that the patient is a football player. This is due to tendency in human reasoning, which they called the **representativeness heuristic**.

> **The representativeness heuristic** is our tendency to misjudge the probability that, for example, not knowing anything about the patient other than the fact that he or she has a torn ACL, ***we are likely to overestimate the odds that the person is a football player based on strong association we have between football players and torn ACLs***.

Strong associations of this kind tend to override our rational capacity to determine probabilities. Let's consider an experiment that Tversky

[3] Griffin, L. Y., Agel, J., Albohm, M. J., Arendt, E. A., Dick, R. W., Garrett, W. E., . . . & Wojtys, E. M. (2000). Noncontact anterior cruciate ligament injuries: risk factors and prevention strategies. *Journal of the American Academy of Orthopaedic Surgeons*, 8(3), 141–150.

and Kahneman ran in the early 1980s showing how we tend to go wrong. Subjects were given the following problem:

- A taxi was involved in an accident one night, but it fled the scene before the police could arrive.
- There are two taxi companies in the city, Green Cabs and Blue Cabs.
- 85% of the city's cabs are Green Cabs.
- 15% of the city's cabs are Blue Cabs.
- Harry witnessed the accident and identified the cab involved as Blue.
- The court tested the reliability of the witness under the same weather and lighting conditions as the night of the accident. They were able to tell that Harry was able to identify the blue taxis successfully 80% of the time. He failed to correctly identify the taxi 20% of the time (Tversky and Kahneman, 1982).

Given these facts, what is the probability that the cab involved in the accident was Blue rather than Green? Given what you know about the situation would you bet that Harry guessed correctly? Pause for a moment and consider your answer carefully. It seems difficult to avoid believing that Harry is probably correct. After all, he is pretty good at telling the difference between blue and green cabs, as we've seen he correctly identifies the color of the cab 80% of the time under the conditions that obtained during the accident.

Our tendency to trust Harry under these circumstances is misguided. However, we are not alone, as Tversky and Kahneman's results showed, most people answered with probabilities over 50%, and many gave answers over 80%. It turns out that the actual probability that Harry correctly identified the cab as blue is 41%. We can discover the actual answer using a relatively simple recipe known as Bayes' theorem (discussed below). In cases like this, we are distracted by the fact that Harry normally does very well identifying the color of the taxi. His success obscures the fact that there are actually very few blue cabs relative to Green cabs in the city.

The low percentage of blue cabs makes a big difference. For example, the chances that Harry is actually in the presence of the blue cab is only 15%. Imagine that he never made a mistake when he identified the difference between blue and green taxis. In that case, there would be a 15% chance that he would correctly identify a blue taxi. However, we know that he correctly identifies the caller of the taxi only 80% of the time. Therefore, this reduces the chance that he would identify a blue taxi correctly from 15 to 12% (15% multiplied by 80%).

It is also important to recognize that a large proportion of cabs in this city are green. Given that Harry correctly identifies blue cars 80% of the time, this means that 20% of the time he is calling green cars blue. Given that most of the cars in the city are Green (85%) this means that there is a 17% chance that he incorrectly identifies a green cab as blue (85% multiplied by 20% equals 17%).

There is a 29% chance (17% +12%) that Harry will identify a cab in the city as blue. He does so correctly 12% of the time and he does so incorrectly 17% of the time. Therefore, most of the time, Harry is wrong when he claims to have seen a blue cab. At this point, you can see how important it is to keep in mind the relatively small number of blue cabs in the city. 12/29 times, when Harry identifies a car as being blue, it is actually blue. This means that he has only a 41% success rate.

9.5 Bayes' Rule

We are regularly bombarded by news and rumors from a variety of sources. When we hear a story, we must decide whether to believe it or not. We do so by weighing a variety of considerations, but essentially there are two factors that are in play:

(1) The plausibility of the story; how likely it is to be true given what we already know?
(2) The reliability of the source of evidence. Should we believe the source?

Imagine hearing a particularly outlandish story. For example, suppose that you heard a report claiming that a 45-year-old philosophy professor had run a mile in under four minutes. If you know something about the history of athletics, you should be skeptical. You might know, for example, that Roger Bannister was the first man to run a mile in under four minutes back in the 1960s. Could a 45-year-old whose primary occupation involves being in a seated position for large parts of the day have run a mile as quickly as one of the greatest runners in the 1960s?

Probably not. Give what we already know, specifically, given what we know about the history of athletics and what we know about being a middle-aged philosophy professor, we are right to think that the story is implausible.

Now, imagine you heard this implausible story from a highly reliable source. Let's imagine that your mother is an expert on sports and that you know she does not lie to you. You are likely to give the story a little more credence, but you are probably still dubious. Could she have misheard a story on the radio? Perhaps someone is playing a trick on her? You know that she is honest and that she is an expert on sport, so you are not ready to simply dismiss what she says. Perhaps you would seek more evidence, you might search on the Internet, ask others, and so on. After all, it is not impossible that a middle-aged professor could achieve something like this. Maybe the professor has been shirking his scholarly and teaching duties to train extensively, or maybe he has some extraordinary genetic gift. The fact that your mother is the source of the implausible story should lead you to lend it more credence than if, for example, your notoriously unreliable cousin told you the same story. An implausible story from an unreliable source should be granted virtually no credence. Now, imagine that your unreliable cousin tells you something that is highly likely to be true given what you already know. Your credence is relatively high even though your cousin tends to be an unreliable source of information.

As we evaluate evidence, we face the challenge of balancing the plausibility of new information, the reliability of its sources, determining how well the new information fits with what we already know, and so on. It might be that new information is derived from multiple sources, none of which is perfectly reliable. How are we to proceed in our deliberations?

When trying to decide how we ought to update our beliefs in light of new evidence a famous equation can be shown to capture the right way to proceed. Known as Bayes' Rule, it is written as follows:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B \mid A) \times P(A) + P(B \mid -A) \times P(-A)}$$

In this equation, A stands for the hypothesis that we are concerned with, and B stands for the new evidence. $P(A|B)$ is the conditional probability that A is true given that B. Another way to say this is that Bayes' Rule tells us how we should adjust our confidence in our initial hypothesis given that we have new evidence.

Bayes' rule is rarely easy to apply, but applying it is less important than understanding the general principles that the rule expresses. Once we understand how its pieces fit together, we can see that it captures in formal terms the right way of updating our initial beliefs in light of new evidence.

In this context, $P(A)$ is what we call the ***prior probability*** that A, which simply means our belief concerning the probability of A prior to receiving the piece of evidence B. Figuring out our priors precisely is not normally achievable in ordinary life. For example, I am not sure of the precise prior probability I would assign to the claim that a middle-aged professor ran a mile in under four minutes. I would assign it a very low probability of being true, but what the number is precisely, I simply do not know.

While we might not know with precision the prior probability of some proposition, we can think of it as the rough likelihood we would assign it given what we knew prior to new evidence. What we hope to learn via Bayes' rule is the probability that A is true in light of the new evidence that B. In other words, we hope to learn the conditional probability of A given B. Bayes' rule shows us how to calculate this. As such, it tells us how our degree of belief in A should be changed in light of the new evidence B.

Reading the equation can be a little daunting for math phobic readers, but it helps to go slowly and break it into its constituent parts. The numerator (the top part of the fraction) is $P(B|A) \times P(A)$. Recall from Section 9.3.2 that when we multiply the probabilities of events we are calculating the probability that both events happen. In the numerator, we are calculating the product of the conditional probability $B|A$ and the probability of A. It should be interpreted as follows: Multiply the conditional probability of the new evidence B given that A is true by the probability that A is true. The result is the likelihood that both $B|A$ and A are true.

It is slightly more difficult to grasp the meaning of the denominator (the bottom part of the fraction). Once again we see

$P(B \mid A) \times P(A)$

on the left. Then we add the result to the value on the right hand side of the "+" namely:

> the conditional probability that *B* given that *A* is false multiplied by the probability that *A* is false.

$P(B \mid -A\ ) \times P\ (-A)$

What the denominator allows is a way of weighing how likely the new evidence is given the truth of *A* and the falsity of *A*; how likely would our piece of evidence *B* have been if *A* were true and how likely would it have been that we would have our piece of evidence *B* if *A* were false.

Let's take an example similar to the kinds of base rate fallacy cases that we saw above. Using real numbers, let's consider a medical test like the pap smear once again. We are trying to learn what the probability is of being sick with the disease given a positive result on the test. Bayes' rule can be set up using *sick* to stand for sick with the disease and *pos* standing for getting a positive result on the test.

We are asking for the conditional probability of being sick given the positive result on the test.

$$P(sick \mid pos) = \frac{P(pos \mid sick) \times P(sick)}{P(pos \mid sick) \times P(sick) + P(pos \mid not\ sick) \times P(not\ sick)}$$

The three values that we need to plug into the equation are the following:

The rate of false positive results

> What is the probability of testing positive given that the subject is not sick (a false positive)? $P(pos \mid not\ sick)$

The rate at which the test catches the disease

> What is the probability of testing positive given that the subject is sick? $P(pos \mid sick)$

What is the proportion of the general population that has the disease?

> What is the prior probability of being sick independently of any evidence provided by the test?
> $P(sick)$

While we are unlikely to actually plug numbers into the equation as we make our decisions, knowing about Bayes' rule is useful in itself insofar as it sensitizes us to the importance of base rates and reminds us of the power of bias in reasoning about probability. When introducing Bayes' rule it is very common for authors to use the example of medical testing because these cases illustrate the difficulty of overcoming bias in decision making. In the medical case, the challenge for us is to overcome the strong association that we feel between a positive test for some disease and having the disease. Knowing about Bayes' rule means resisting the automatic pull of strong associations like these.

Bayes' rule helps us to see that while the positive test result certainly counts as new evidence, it should not be understood to function like an on–off switch. Just because you have evidence that you have a disease, for example, this evidence cannot be understood apart from the larger context of what we already believe. New evidence should cause us to update our confidence in our beliefs according to the rules of probability theory and with a full understanding of its significance in relation to the relevant base rates.

10 The Informal Fallacies

Formal fallacies of the kind we studied in previous chapters involve failures of reasoning that we can identify by reference to a standard set of logical or probabilistic errors. Many of these are straightforward kinds of mistakes that we can easily learn to spot with a little education in the relevant formal methods. As we saw above, these errors are sometimes due to innate cognitive tendencies or heuristics that cloud our judgment under certain circumstances. Knowledge of where we are likely to go wrong combined with knowledge of basic formal methods can save us from falling for formal fallacies. Unfortunately, not all fallacies follow an obvious formal pattern. In this chapter, we turn our attention to the more challenging realm of the informal fallacies.

The so-called informal fallacies are not simply mistaken patterns or violations of formal rules. Instead, the fallacies we consider in this chapter are more like impediments to successful inquiry. One way to understand them is as failures to exhibit the epistemic virtues we discussed in earlier chapters. These fallacies result from our weaknesses; they work by distracting us, by manipulating our emotions, by exploiting our laziness, or by playing on deeply engrained habits of thought, strong associations, or prejudices. Some fallacies introduce premises that are irrelevant to the matter under consideration, some rely on faulty assumptions concerning the nature of the problem or they introduce inappropriate methods for deciding a question.

As mentioned in previous chapters, there are an infinite variety of ways that reasoning can go wrong. Our discussion will touch on the most prominent of the informal fallacies. But that should be enough to sensitize you to the kinds of problematic patterns of reasoning that we find commonly in political life and commercial speech.

10.1 Distractions: Straw Man Arguments, Ad Hominem, Tu Quoque, Red Herrings

One way that fallacious arguments impede clear thinking and decision making is by distracting us. Misdirection is a familiar part of the rhetoric of argumentation. In a competitive context, rather than objectively presenting the strengths and weaknesses of one's argument, debaters, attorneys, or salespeople tend to overemphasize the strengths of a favored position and minimize its weaknesses. There are a variety of illegitimate tactics whereby one can direct an audience's attention to those aspects of an argument that make the proponent's own position seem stronger and their opponent's weaker.

10.1.1 Straw man fallacy

If I accurately present some claim, I might not be able to persuade you to reject it. However, if I can make that claim look less plausible than it actually is, I may be able to persuade you to reject this weaker version of claim. If you are tricked into thinking that the weaker version of the claim is identical to my opponent's claim, then you will think that you have rejected my opponent's claim and that I have won. By adopting this dishonest strategy, I am committing the straw man fallacy.

Arguing against a weaker version of your opponent's position is like physically triumphing over a scarecrow dressed in your enemy's clothing. You can easily defeat a straw man, but you have not defeated your real opponent. Your rhetorical strategy works only insofar as you can trick your audience into mistaking the straw man for your opponent.

For example, imagine arguing against libertarian political ideals. Libertarians generally place a great deal of importance on protecting the rights of individual persons and are suspicious of collectivist projects that subordinate individual interests to the interests of groups. You would be committing a straw man fallacy if you suggested that

Libertarians deny the value of community

Here, the straw man version of libertarianism is so extreme that it fails to stand the test of serious scrutiny. It would be easy to refute someone who denied the value of community. One could do this simply by showing that community life has some value. If libertarianism necessarily involves the denial of the value of community, then it becomes a relatively implausible position to take. This is an example of the straw man fallacy insofar as it shifts the audience's attention from the libertarian commitment to individual rights to the libertarian suspicion of collectivist political ideologies. The fallacy is to replace that suspicion with the claim that libertarians deny the value of community. This goes beyond simply being a caricature of libertarianism by asserting a false claim about Libertarians, namely that they deny the value of community. It is difficult to imagine any reasonable position in political philosophy that is committed to the denial of the value of community. Depicting libertarians as such is to present a straw man in their place.

The straw man fallacy is pervasive in contemporary political discourse. It generally consists of replacing one's opponents' position with some outrageous or extreme version of that position. To take another example:

The animal rights lobby wants to prevent us from having pets.

It might be the case that there are members of the animal rights movement who are against the practice of keeping pets. However, it would be unfair to present this as a defining characteristic of the entire movement. Similarly, while it might be the case that some supporters of President Trump are anti-Semites, it would be a mistake to take extreme or unusual instances of Trump's supporters as representative of the movement that supported his election.

Similarly, if we were to characterize scientists whose research involves animal experimentation in the following way:

Vivisectionists are bloodthirsty torturers who enjoy hurting innocent creatures.

We would also be falling into a kind of straw man argument. It might be the case that there are some scientific researchers who enjoy hurting animals. However, winning the argument against bloodthirsty torturers who enjoy hurting innocent creatures is not a victory against scientific research that involves the use of animals. By refocusing an audience's attention with emotion or strong language, a dishonest debater can dramatically reduce the strength of their opponent's position. It is easy to win the argument against bloodthirsty torturers, anti-Semites, or those who would take away our pets. However such characterizations of their respective opponents are false.

Another way that straw man arguments operate is by criticizing some relatively trivial aspect of the opponent's argument in a way that exaggerates the weakness of the overall argument. For example:

> *Fred is in favor of the right to bear arms, but he is assuming that it will be possible for everyone to afford a gun. There are many poor people who will not be able to buy a gun. He's really arguing for rich people to have guns.*

It might be true that Fred, in arguing for the right to bear arms had not considered the possibility that some people will not be able to afford a gun. However, the fact that some people will not be able to afford a gun is not relevant to the question of whether people should have the right to own a gun. Perhaps the gun rights advocate would be willing to subsidize guns for the poor, or perhaps there might be ways that guns could be produced more cheaply. In any event, the question of affordability is irrelevant to the central question. The argument for the right to bear arms is not significantly weakened by facts about the affordability of guns.

10.1.2 Ad hominem arguments: arguing against the person rather than against the argument

Ad hominem arguments introduce characteristics of one's opponent that have nothing to do with the matter under consideration in order to distract one's audience from the real substance of the issue, specifically the quality of the argument. Associating some irrelevant characteristics of the opponent (either favorable or unfavorable characteristics) with their thesis is meant to influence the audience illegitimately to reject the thesis. For example, imagine that Sally is arguing for the benefits of international free trade agreements.

Her opponent argues:

> *Of course Sally is in favor of neo-liberal economic policies, she says that free-trade benefits the poor, but I heard that she doesn't give a penny to charity. So much for caring about other people. So called free trade is exploitation pure and simple. . . not that Sally would care about that. Have you noticed how little she tips at restaurants?*

The problem with her opponent's argument is that it focuses on a characteristic of Sally that has no bearing on the benefits or costs of free trade agreements. Whether Sally gives to charity, tips generously, or not is irrelevant to the strengths or weaknesses of her argument in favor of international trade. The harmful effect of this fallacy is that it distracts the audience from her argument and instead creates a sideshow where we are led to impute a lack of generosity as the chief motivation for her argument. Ad hominem arguments work in a variety of ways:

They creating an association between an opponent's claim and some irrelevant feature of the opponent they mislead the audience such that they end up make and something like the following set of inferences:

> *Sally is in favor of free trade agreements because she's greedy.*
>
> *Being greedy is bad, therefore being in favor of free trade agreements is bad.*
>
> *Therefore free trade agreements are bad.*

They incorrectly conflate the reasons she provides for believing the thesis with her moral characteristics:

> *Sally is in favor of free trade agreements because she's greedy.*
>
> *Being greedy is not a good reason to be in favor of free trade agreements*
>
> *Therefore Sally is giving a bad argument for free trade agreements.*

They play on our desire to affiliate with good people and not affiliate with bad people.

Sally is in favor of free trade agreements because she's greedy.

Being greedy is bad.

Bad people are in favor of free trade agreements.

I don't want to be a bad person.

Therefore I am not in favor of free trade agreements.

Ad hominem arguments are fallacious insofar as they distract the audience from the actual issue at hand, through focusing on some irrelevant, but usually emotionally significant detail concerning one's opponent. Ad hominem arguments, which do not make reference to some shameful or emotionally charged feature of one's opponent tend not to work so well. For example, if in response to Sally someone were to say:

> How can you accept Sally's views on free trade? She's under 7 feet in height.

None of us would be moved to reject her position because her height is so obviously irrelevant to the matter under consideration. The kinds of attributes that ad hominem arguments tend to introduce for rhetorical effect are those that move us emotionally. We either feel some revulsion toward the person described and are tricked into associating that revulsion with the person's position, or we identify those characteristics with a group with which we do not wish to be affiliated and are tricked into rejecting that person's position as a way of affirming our preferred affiliation relations.

10.1.2.1 *Tu Quoque* (The "You Too" Fallacy)

This fallacy is a familiar part of ordinary arguments. It basically involves attacking one's opponent in an effort to defend oneself. The "look who's talking" response to an opponent's argument is intended to distract attention from the weakness of one's own position by introducing an irrelevant aspect of one's opponent's character or behavior. For example:

> *The manager accused me of not getting into work on time last Wednesday. But seriously, that guy is on Facebook all day long, he might be in the office, but he's not doing any work.*

Let's assume that it is true that the manager is wasting time on Facebook during work hours. This fact has no bearing on whether or not the proponent was late on Wednesday. Mentioning his manager's timewasting is simply a distraction rather than a genuine defense.

The "you too" fallacy seems compelling insofar as it makes the opponent look like a hypocrite. Just as in the case of the ad hominem argument discussed above, the fact that the opponent is guilty of hypocrisy does not mean that what they are saying is false. Their defective character is independent of the quality of their argument.

10.1.3 Red herrings

A red herring is a deliberately misleading diversion that is introduced in order to distract the audience from the actual matter under consideration. For example, it can be a plausible new line of reasoning that is related somehow to the subject matter of the original argument but which is ultimately irrelevant to the central issue.

> *Drago: We should not allow internet service providers to sell the records of our activity on the internet, it is a violation of our privacy.*
>
> *Fritz: Surely you realize that the right to privacy is not protected in the constitution, it is a much more recent invention.*
>
> *Drago: But isn't the Roe vs. Wade decision of the supreme court based on their view that there is a constitutional protection of privacy?*
>
> *Fritz: Yes, you're right, but I think that raises real questions about the constitutional status of legalized abortion.*

The original argument is concerned with privacy, but not with whether there is a constitutional protection for privacy. Fritz manages to distract Drago by leading him into a discussion of whether there is a right to privacy in the constitution. The discussion of the constitution, *Roe v. Wade*, abortion, and so on is a red herring.

Cleverly deploying a red herring is a way of changing the subject or leading your audience off the trail in ways that can distract them from the weakness of your position. It is common for people who are found guilty of some infraction to introduce red herrings to distract from their guilt. Consider the following:

> *The professor claims that I plagiarized my paper, but this is completely unfair. Everyone knows there is no such thing as a genuinely original essay. She should spend her time helping us to understand the material properly rather than playing detective with our papers.*

The issue at hand was, presumably the fairness or unfairness of the plagiarism charge. The student's attempt to change the topic involves introducing a new argument, namely that the professor is not doing her job and should be doing something other than enforcing plagiarism rules. Notice that the student is less concerned about convincing his audience that the professor is fulfilling her duty and more with making sure that the audience forgets the original issue (the fairness of the plagiarism charge) focusing instead on the argument over the professor's duty.

It is often the case that the red herring is introduced with the full knowledge that it will provide a tempting topic for one's audience to pursue. In this case, the audience will be tempted to consider whether the professor really should be spending time "playing detective". Is the professor deriving some kind of perverse satisfaction in catching plagiarists? Perhaps the professor is inappropriately using her position to persecute innocent students instead of performing her duties correctly. . . Once the audience becomes interested in juicy topics like this, the red herring has successfully distracted attention away from the student's guilt or innocence.

10.2 Appeals to Emotion, Appeals to Popularity and Tradition, the Genetic Fallacy

It is very common for advertisers, politicians, and the like to use emotion as a way of persuading their audience to hold some set of beliefs. If there is some strong positive emotional connotation to a belief or if such a strong positive connotation can be created in the minds of an audience, they will tend to assent to that belief. Similarly, an audience can be moved by shock or distress to accept some claim. Consider, for example, the use of shocking or graphic photographs by animal rights advocates or by anti-abortion activists. The goal of such activists is not to rationally persuade their audience. Instead, they hope to leverage our visceral reaction to the images rather than providing a rational argument.

The tactic of using emotion to undermine the ability of one's audience to think clearly is manipulative and often deceptive. One could, for example, make a movie in which many perfectly innocuous and morally neutral activities look disgusting and shocking given the right kinds of camera angles, music, and so on.

We should be careful not to allow our sense of disgust or shock to guide our moral judgments. A mother spanking her toddler would be a shocking sight to residents of Cambridge, Massachusetts, but not necessarily to Kansans. Judgment concerning what does and does not count as disgusting or shocking vary with geography, historical location, culture, and so on. However, the fact that a wide variety of divergent emotional reactions are elicited by the same phenomena does not mean that there is no matter of fact as to whether some act is right or wrong. It only shows that emotions are not our best guide to learning what is right or wrong. Whether, for example, spanking a child, interracial romantic relationships, or facial tattoos, are morally problematic should not be decided by reference to the emotional reactions that they elicit.

As a child I remember wanting to believe in Santa Claus long after I had good evidence against his existence. The belief that there is a generous magical being who brings presents on Christmas Eve is extremely pleasant. I recall that for some period of time, I believed (or at least I said I believed) in Santa Claus only because I enjoyed believing in Santa Claus.

As we saw above, we could also assent to some belief because of negative emotions like fear, vanity, disgust, spite, and the like. If, for example, one finds some sexual practice disgusting or unpleasant, one's emotional reaction should not necessarily guide one's moral judgments. There may be good reasons to condemn that practice, but my emotional reaction to it is not, by itself sufficient justification.

Sometimes, a tangle of emotions is in play, overriding our rational capacity and making it difficult for us to see clearly. Think of the following combination of loyalty and fear motivating the following statement:

I can't imagine that my son Billy would ever do something like that, just the thought of it would break my heart.

This parent is denying that Billy did what he is accused of doing simply because the idea of Billy doing that thing is distressing. However, the emotions involved are complicated and emotions like loyalty are noble in some contexts. However, if we are interested in the pursuit of truth, we will be on our guard against the distraction posed by strong emotion.

The appeal to emotion, whether it is an appeal to fear, pity, love, or some other feeling is a fallacious tactic to deploy on an audience. It is also important that when we are in the position of the audience, we not allow emotions to override our deliberative capacity. Other fallacies that have been identified by philosophers over the years involve versions of this kind of fallacy. For example, appeals to pity, arguments from force, or the use of threats to win arguments all have the same basic flaw that we have seen here.

10.2.1 Appeals to popularity or tradition

The fact that some claim is popular or has been accepted traditionally by some group is usually irrelevant in an argument. For example, it is fallacious to argue:

Acupuncture has been used for centuries in Chinese medicine therefore it must have some health benefits.

In traditional societies, people went to sleep soon after it got dark, so the practice of staying up late in modern society is unhealthy.

None of my friends put their sponges in the dishwasher, I don't think it's a good idea.

We always go to my mother's house for Thanksgiving, it makes no sense to go to Jamaica.

What's the point of going to the opera? It's the 21st century, nobody listens to that stuff.

In each case, the appeal to tradition or popularity is irrelevant to the claim under consideration. The fact that acupuncture has been used for centuries does not directly bear on the question as to whether acupuncture is effective. Many medically ineffective and even harmful practices have persisted for long periods of time. Few of us today would defend bloodletting as a medical practice in spite of its long history in medieval Europe.

Likewise, the fact that some claim is popular has no bearing on its truth value. Popular opinion by itself is not trustworthy. With respect to many matters the majority of people hold false beliefs. Sugar does not cause children to be hyperactive, the moon does not have a permanently dark side, and cracking your knuckles will not give you arthritis even though most people believe these things.

In the case of the sponges in the dishwasher, it's also possible that the sample of friends, which the speaker draws upon, might not be a reliable one. Perhaps all of this person's friends are foolish or poorly informed. The popularity of a claim among such a group should not encourage us to follow their lead. However, while the popularity of a claim by itself is irrelevant, if this person's friends were experts on dishwashers and germs then their views would be relevant to the question of whether or not to put sponges in the dishwasher.

Notice that the fallacy in question arises when the appeal is to popularity or tradition by themselves, rather than say the popularity of a claim among some authoritative group of experts.

10.2.2 The genetic fallacy

It is a fallacy to judge an idea or claim true or false based on irrelevant historical associations that it might have. The truth or falsity of some claim is usually independent of the person who happened to discover the claim, the kinds of people who have accepted this claim in the past, and the manner in which the claim was discovered. Therefore, for example, if some medical claim was discovered using unethical means, this, by itself, is irrelevant to the truth or falsity of the claim. Similarly, we can imagine an important scientific breakthrough like the discovery of penicillin taking place as the result of an accident. The fact that the discovery happened by chance is ultimately irrelevant to the merits of the scientific claims being made.

As we saw previously in our discussion of Bayes' rule, *it is rational* to take the reliability of a source into consideration when deciding how much credence to give to evidence that the source provides. However, as we also saw, the fact that a source of evidence is not highly reliable does not mean that the evidence it provides is *necessarily* false. Some source of evidence might be so unreliable that it is rational to assign it such low levels of credence that you decide it is not worth considering with much care. You might decide that you have finite time or other resources to allocate and that these are better spent on other investigations or inquiry. This decision means that you will more or less ignore that source of evidence. However, it is important that this allocation of resources be distinguished from the claim that the evidence provided by that source is false. To assume that it is *necessarily* false is fallacious.

The genetic fallacy is tricky insofar as it is easy to confuse the basis for an inference concerning the truth or falsity of a belief with the judgment concerning the degree of credence we should assign to some proposition. If we make an inference from the unreliability of a source of evidence to its falsity, we are making a mistake. If we judge that a piece of evidence coming from a source of evidence that we find unreliable is to be given less weight than a piece of evidence coming from a more reliable source we are being rational in our evaluation of the evidence.

One way to think of the genetic fallacy is as follows: Just because someone or something has unsavory origins does not mean that they or they necessarily must be dismissed, shunned, or ignored. For example, the fact that your parents were criminals does not mean that you are a criminal or that your character should automatically be assessed in light of your parents' crimes. However, perhaps it is the case that the children of criminals are more likely to become criminals than the children of noncriminals. If we knew that they were statistically more likely to become a criminal, it does not mean that they necessarily became a criminal.

10.3 Informal Fallacies Related to Causal Judgments

Causation poses deep and difficult philosophical problems that are well beyond the scope of this book. We all have some pre-theoretical sense for what it means to claim that some event causes another event. The idea of cause is central to our reasoning about the world. It figures in our judgments concerning what is and is not relevant to our goals and actions. Causes figure, not only in our plans for the future but also in our explanations of what has happened in the past. When we say that the thrown stone caused the window to break, that consuming alcohol causes intoxication, that smoking causes cancer, that cracking one's knuckles causes arthritis, that the car's door was scratched by an angry student, that we are restricting the dog's diet to prevent him from getting too fat, and so on, we are engaging in causal reasoning of various kinds. It is not necessary for us to fully understand the nature of causation itself in order for us to recognize that some judgments concerning causal relations are simply fallacious. We recognize, for instance, that just because one event follows after another does not mean that the first event caused the second. Likewise, we understand that the fact that two events are correlated does not, by itself, entail that they are causally related.

10.3.1 After this therefore because of this (Post hoc ergo propter hoc)

This fallacious pattern of reasoning takes the fact that event B happens after event A to license the judgment that A caused B.

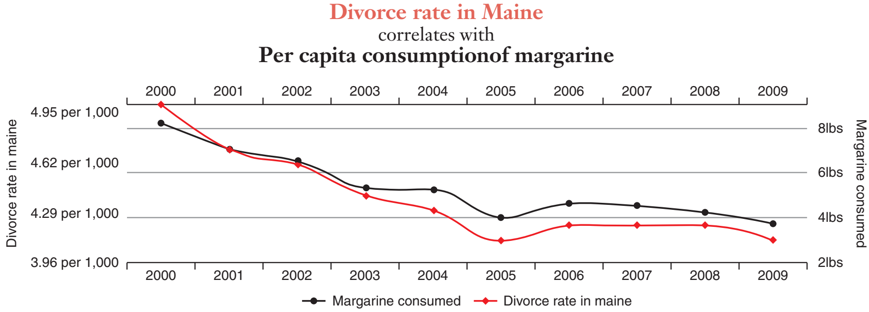
> A occurs before B.
>
> Therefore
>
> A is the cause of B

The simple fact that one event precedes another does not mean that the two events are causally related. I might desperately seek some causal explanation for why I was unsuccessful in a job interview. It might occur to me that evening that I brushed my teeth with a new brand of toothpaste that morning. It is tempting to think that there is some causal significance to the fact that this was the first day I tried the new toothpaste and then, later that day, failed to get the job. If I ascribe causal significance based solely on the fact that the toothpaste event preceded the job interview event, I am committing the post-hoc fallacy.

Note that I might turn out to be correct about the toothpaste. Perhaps, on further investigation, I learn that the toothpaste had some effect on my cognitive capacities and that this caused me to perform badly in the interview. Given this new information, I am entitled to claim that the new toothpaste caused my bad performance. By contrast, if all that grounds my judgment of a causal relation between the toothpaste and the interview is their order in time, my reasoning is flawed.

10.3.2 *With this therefore because of this (Cum hoc ergo propter hoc)*

We frequently hear that correlation does not equal causation. It is true that there are coincidental correlations that do not indicate any causal relation whatsoever. Tyler Vigen has a wonderful website (http://www.tylervigen.com) devoted to collecting spurious correlations. Vigen notes for instance that there is a 99.26% correlation between margarine consumption and divorce rates in the state of Maine between the years 2000 and 2009. If we were tempted to conclude that margarine consumption must be causally related to divorce rates in Maine, we would be committing a fallacy.



The fallacy of Cum Hoc Ergo Propter Hoc (*with this therefore because of this*) is equivalent to the inference that since two events, A and B happen together, they are therefore causally related. As in the case of the post-hoc fallacy, the problem is that two events merely being correlated or following after one another does not entail the existence of a causal relation. In the case of Cum Hoc Ergo Propter Hoc, it might be the case that the correlation of A and B are coincidental.

Alternatively, it might be case that there is some third factor doing the causal work behind the scenes. Let's say that we notice a strong correlation between buying cough medicine and buying throat lozenges. Clearly, we would not be tempted to think that buying cough medicine causes people to buy throat lozenges, or vice versa. Instead, we would say that a third factor, that is, having a cold, for example, is the cause of people buying both cough medicine and throat lozenges.

Screening off these third factors is not always such an obvious matter. Take, for example, the correlation between majoring in philosophy in college and having a high median mid-career income. It is a striking fact that by this measure, philosophy majors financially outperform most college majors (with the exception of highly formal disciplines like math, engineering, economics, physics, etc.). Those of us who teach philosophy are proud to point to reports in the *Wall Street Journal* and elsewhere documenting the striking level of financial success associated with the study of philosophy. But is our pride justified? Does studying philosophy really cause students to have more financial success than studying biology, history, or psychology?

Before we become too excited about the benefits of studying philosophy, we should look to other possible explanations of the correlation between high median mid-career income and majoring in philosophy. Perhaps, for instance, the fact that a higher proportion of male students pursue the major in U.S. colleges and universities, along with the fact that American men tend to earn more than American women is enough to explain the dramatic difference between, for example, philosophy and psychology majors. Controlling for the proportion of men and women in the sample might make the effect disappear. If this were the case, then the high salaries of philosophy majors would be due to this other factor (the disproportionately male population of philosophy students in the United States). Alternatively, there might be what is called a **selection bias**, such that intellectually curious, and highly confident, or argumentative students might be drawn to philosophy. If the pool of students who decide to pursue philosophy are already intellectually stronger in some dimension than the students who decide to pursue biology or history, then this third explanation would further also the argument that studying philosophy causes higher mid-career earnings.

Until these alternative (third factor) causes are eliminated, it is wise for philosophers to be cautious when they claim a causal role in the high earnings of their students.

10.4 Informal Fallacies Related to Misunderstanding the Burden of Proof

By the "burden of proof" what we mean is the responsibility to defend or justify some claim. If, for example, you are minding your own business and I ask you to accept some unusual claim, it is my responsibility to provide good reasons for you to accept the claim. It is not your responsibility to convince me of why you should not accept my claim. The burden of proof is on me to convince you, it is not your job to convince me that the claim is false.

> In legal contexts, *the presumption of innocence*; the idea that a defendant is innocent until proven guilty, is equivalent to the principle that *the burden to prove a defendant guilty falls on the prosecutor.* It is not the defendant's responsibility to prove himself or herself innocent beyond a reasonable doubt.

Some bad patterns of argument involve concealing or shifting the burden of proof in order to make the proponent's thesis look more plausible. In this section, we will examine two fallacies of this kind: The argument from ignorance and the fallacy of begging the question.

10.4.1 The argument from ignorance

Imagine sitting around the campfire with your friends. As often happens during late night sessions like this, someone begins to tell a story of the supernatural. Perhaps one of your friends describes his experience of seeing a ghost. He tells the story with an earnestness and level of detail that lead you to think that maybe ghosts are real after all. Perhaps one of your friends is less gullible and denies that your friend really saw a ghost. At moments like this the storyteller often replies in the following way:

> How can you be so sure that I didn't see a ghost? You can't prove that I didn't see the ghost. Don't be so closed minded.

The storyteller is beginning to commit the fallacy of the argument from ignorance. This fallacy is often deployed in support of supernatural or outlandish claims. It takes the following form:

> You can't prove that it's not the case that X
> Therefore X is the case.

The trouble with this line of argument is two-fold. On the one hand, my lack of knowledge with respect to the topic of ghosts or goblins is irrelevant to the claim that ghosts or goblins exist. The fact that I have no real expertise concerning the moons of Jupiter is similarly inconsequential with respect to whether or not there is life underneath the ice on some moon of Jupiter. Imagine a scientist claiming that there is life underneath the ice on some moon of Jupiter because John Symons is unable to prove that there is no life on the moons of Jupiter. The fact that John Symons knows almost nothing about astrobiology does not serve as evidence for or against the claim that there is life on Jupiter's moons.

The fallacy of arguing from ignorance violates a basic norm with respect to justification and evidence. It is constitutes an illegitimate attempt to confuse one's audience with respect to the burden of proof in some discussion. The burden of proof falls on the proponent of the argument to make their case. The burden does not fall on the audience to disprove the proponent's thesis.

10.4.2 Begging the question

The fallacy of begging the question is another illegitimate way of confusing one's audience with respect to what needs to be proven and how evidence in support of a conclusion should be provided. Examples of this fallacy involve simply asserting the conclusion in a way that makes it appear to the audience that the conclusion is derived from some premises rather than being a simply asserted in some form as a premise in the argument. Begging the question involves the use of what is called **circular reasoning**. Circular reasoning is problematic insofar as it illicitly assumes the conclusion in the premises of the argument. Therefore, for example, if I were to claim that:

> I know Harry is honest because he always tells the truth
>
> and
>
> he swore to me that he was honest.

I could be asked why I believe that he is telling the truth about being honest. If my belief that he is telling the truth rests on my belief that he is honest, I am engaged in an obvious instance of circular reasoning.

Begging the question is the fallacious use of an assertion that is the same as or equivalent to the main contention of the argument as a reason for accepting its main contention. Thus, begging the question involves simply restating the contention in different words and hoping that the audience does not recognize what the proponent is up to.

> Why does everyone love Dustin?
> Because he is the most popular kid in the school.
>
> How do you know that God exists?
> Because of divine revelation.

More subtle examples of begging the question make it more difficult to see that the conclusion is being restated in the premises. For example:

> Why don't you make the boys wash the dishes Dad?
> Because kitchen tasks are women's work.
>
> Abortion is wrong because it's murder.

Simply asserting what is supposedly being proved does not count as a proof. The fallacy of begging the question is nothing more than an unsupported assertion of the conclusion of the argument in a deceptive manner. As such, it relies on confusing the audience about what needs to be proved.

Unfortunately in ordinary contemporary discourse, especially among journalists and television commentators the phrase "begs the question" is almost always misused. In most cases if they say that some claim or state of affairs begs the question, they are simply saying that it makes them want to ask a question. Such people embarrass themselves by attempting to appear clever or educated by using a phrase that they misunderstand because it has a scholarly halo around it. However, the misuse has become so common that it is becoming increasingly difficult to say that it is an error.

10.5 True Beliefs Can Be Supported by Bad Arguments

You can be led to a true belief for reasons that really shouldn't lead you to that belief. Therefore, for example, I have a Swedish friend who believes that babies sleep better in the cold than in a warm cozy crib. She tells me that Swedish children are often left outside in their strollers to nap while parents eat their lunch or do their shopping indoors. The fact that this is a common practice, or that it is traditional in Scandinavian countries to leave napping babies outside in the cold is not itself a good reason to believe that babies sleep better in the cold. It is not good policy to accept some view just because people have traditionally believed it or because it's a widely accepted belief. However, it might be the case that babies should sleep outdoors in cold weather. Perhaps at some point there will be some scientific evidence supporting her claim. Even if such evidence is forthcoming, it is still the case that her current belief is not supported by good argument.

Take another example, let's say I claim that if I run a marathon then I'll be tired. You notice later that I am tired. You conclude on the basis of this that I must have run a marathon. On the face of it, this is clearly a fallacious line of reasoning. This is the fallacy of affirming the consequent that we saw in Chapter 8.

Given that you know:

If John runs a marathon then John will be tired.

And

John is tired.

You should not conclude that John ran a marathon.

To understand why not just consider other ways that I could have become tired. Clearly, the fact that running a marathon would make me tired does not preclude the possibility that I could have stayed up all night studying, that I could have run a half marathon, that I could be an insomniac, and so on. There are countless ways that could have become tired. Moreover, the fact that I am currently tired should not automatically lead you to believe that I ran a marathon. However, it is possible that in fact I did run a marathon! If that were the case, then you would have arrived at a true belief for bad reasons.

In cases like this, situations where we hold true beliefs as a result of faulty or fallacious lines of reasoning we ought to deny that the person who holds the true belief has genuine knowledge. In these cases, it's like having knowledge by accident. Therefore, if I correctly guessed the population of Kinshasa, we would not be inclined to say that I have genuine knowledge. Similarly, if I come to some belief that *A* via a faulty line of reasoning and even if it turns out to be true that *A*, it makes sense to deny that we really know *A*.

11 Understanding Formalism: Sentential Logic

Chapters 7–9 have been devoted to explaining the ways in which arguments tend to go awry. As we have seen, by abstracting from the content of our sentences and focusing instead on their form, we can overcome the biasing effects of content and can develop our sensitivity to common failures in reasoning. We can thereby avoid them in our own thinking and decision making. We have learned some of the important benefits of formal reasoning and have seen why we must go beyond ordinary common sense reasoning in some important contexts. However, we have not yet explained in detail what a formal method is, and what it can allow us to do.

Formal methods are used extensively in mathematics, computer science, linguistics, engineering, and philosophy. A formal method begins by defining a simple system. Examples of such systems might include axiom systems, computer programming languages, logics, and the like. The point of building an artificial language like a computer programming language, for example, is to make certain that no errors or confusion inadvertently infect some important purpose. For practical purposes, for example, errors in software must be kept to a minimum. Building an artificial language from nothing means carefully introducing the parts of the language in ways that are clear and unambiguous. We want all parts of our system free from obscurity and fuzziness. Our task in constructing an artificial language is not to be faithful to the gloriously fuzzy and contextually sensitive reality of natural languages. Instead, we are building something whose parts we understand and whose behavior is constrained by rules we fully understand. While we may be surprised by where these systems take us, we will know exactly how they took us there.

In building an artificial language, one defines what it means to be a sentence of that language, what its syntactical rules are, and what the properties of those sentences are. The definitions that we provide are formal, in the sense of being precise, mathematical characterizations of what it is that we are talking about. Nothing other than features that can be specified precisely will be included in the definition of the formal language. Obviously, a formal language differs significantly from a living natural language. We do not have to worry about the fluid meanings of words; the unstable grammar; the influence of context, history, and culture; and the like.

In this chapter, we introduce a simple formalism called sentential logic. Sentential logic is also known as *truth-functional logic* or *propositional logic* (for reasons, we will see soon). Sentential logic is easy to master and, in itself, it is not a particularly complicated mathematical formalism. In spite of its simplicity, sentential logic has deeply interesting properties and holds great interest for those of us interested in argument and reasoning.

Sentential logic gives us a method for analyzing the formal features of arguments that are related to combinations of simple declarative sentences. Specifically, it is focused on the ways in which declarative sentences can be formed into more complicated declarative sentences using the logical connector words "and," "or," "if. . . then," and "not."

It is useful to gain some acquaintance with sentential logic insofar as it offers us a way of thinking formally and systematically about arguments and proofs. As we examine how to use and think about the properties of sentential logic, we will gain insights into the nature of proof and validity more generally. We learn what it means to prove that an argument is valid, and we learn about what kinds of things count as proofs. We get a sense for the variety of ways to give proofs, and we begin to learn what counts as a good proof.

Once one grasps the basics of this formalism, other formal methods become much easier to appreciate and understand. In this way, readers who approach mathematics and computer science with trepidation will find sentential logic to be a friendly first step into formal reasoning.

11.1 Declarative Sentences, the Building Blocks of Arguments

The starting point for sentential logic is the simple declarative sentences and its properties. It helps us to understand how these properties change when declarative sentences are combined or broken apart in various ways. Since arguments are composed of strings of sentences, some properties of arguments can be usefully understood using sentential logic. As we shall see, sentential logic is just the starting point for the study of the formal properties of arguments, but it is a necessary and simple first step. The techniques we will learn studying sentential logic will serve us well as we proceed more deeply in the study of formal methods.

Let's begin with the building blocks of sentential logic, the declarative sentence: A declarative sentence, as we saw in previous chapters, is a sentence that makes a claim about the way things are. We can contrast declarative statements with questions, instructions, involuntary noises, exclamations, etc. Usually, we say that a declarative sentence is either true or false. By contrast, a question or a command is not usually understood to be either true or false. For example,

(a) Las Cruces is north of Santa Fe

is a declarative sentence which happens to be false. While

(b) Some mammals lay eggs

is true (because of the platypus). Whereas

(c) Make me a sandwich.

or

(d) Ouch!

or

(e) Where is the Portuguese class?

are not declarative sentences. We say that (a) and (b) have a *truth-value*, while (c), (d), and (e) do not. In classical logic, we restrict ourselves to declarative sentences that are either true or false. Notice that this restriction will exclude many apparently declarative sentences. Take, for example, a self-referential sentence like

This sentence is false.

The truth-value of this sentence poses a conceptual challenge. After all, if it is true, then it is false. But, if it is false, then it is true. Sentences of this kind are known as paradoxes. The paradoxical quality of the sentence is not due solely to the fact that the sentence is making reference to itself. Self-reference is not, in itself, paradoxical. For example, consider a sentence like

This sentence has five words.

which refers to itself and is straightforwardly true. Some philosophers have worried that our usual starting point in logic treats only declarative sentences with only one of two truth-values (true or false) is too restrictive. They have complained that the focus on declarative sentences of a particular kind unnecessarily excludes paradoxical and self-referential statements. Other philosophers have challenged the so-called **principle of bivalence** which states that every declarative sentence has exactly one truth-value, either true or false. These philosophers have a point. The focus on declarative sentences and adherence to the principle of bivalence involves a restriction on the kind of subject matter under consideration. But recall that we are building a formal system and that our purpose at this stage is not to be faithful to the way ordinary language or ordinary usage operates. In building our formalism, we are entitled to ignore nuance and subtlety for the sake of clarity and control.

However, it is important to be conscious that taking the declarative sentence as the basic building block of sentential logic and accepting the principle of bivalence involves a decision. Other possible formalisms are possible. There are logics with more than two truth-values and logics in which the principle of bivalence does not hold. Sentential logic is one among many formalisms; we make no claim that it is the unique and correct logic governing arguments. In fact, we will learn another formalism—first-order logic—in later chapters. As we shall see, this formalism carves up ordinary sentences in natural language in ways that differ significantly from sentential logic.

Declarative sentences: For example, "Fabiola is Canadian" are understood as having the values *true* or *false*. "True" and "false" are the only values that we assign to sentences in sentential logic. It is worth noting that, for the purposes of our formalism, we need not worry about the philosophically problematic nature of words like "truth." In fact, we could replace "true" and "false" with alternative binary values "1" and "0," "ON" and "OFF," or some other less philosophically significant terminology. While truth is an important concept, for the purposes of the formalism of sentential logic, we need only concern ourselves with the stipulation that declarative sentences have a binary value.

11.1.1 Variables in sentential logic

Sentential logic abstracts from the meaning or content of sentences in order to focus on the form of arguments. One way it does this is by allowing variables to stand for individual declarative sentences. As we saw in Chapter 4, variables in mathematics are letters that stand in for values or objects of various kinds. In algebra, for example, a variable can represent some unknown or unspecified numerical value in a problem or an equation. If you are told that $2x = 10$ and asked to solve for x, you would know that in this case $x = 3$. In more advanced mathematics, variables can stand for other kinds of objects, not just numbers, but vectors, functions, and matrices. In sentential logic, **variables represent declarative sentences**. Specifically, declarative sentences will be represented in our logic by lower-case letters from the end of the alphabet:

$p, q, r. . .$

These letters are variables, each of which will stand for some declarative sentence.

11.1.2 Declarative sentences and propositions

Basic declarative sentences are sometimes thought to refer to *propositions* and the part of logic which deals with the relations between propositions is sometimes called the *propositional calculus*. In fact, some philosophers would criticize my use of the term "declarative sentence" in this context. They would argue that the word "sentence" refers to the particular marks on the page or a speaker's verbal utterance in a specific case. In logic, they argue, we should be talking about propositions, not sentences. Propositions are what the sentence conveys. Different sentences such as "La nieve es blanca," "*Der Schnee ist weiss*," and "Snow is white" express the same proposition, and it is this proposition that interests us in this context, not the sentences themselves. There is clearly something right about the idea that someone who understands both Spanish and German recognizes that "La nieve es blanca" and "*Der Schnee ist weiss*" both assert the same thing about the world. While the sentences vary, the proposition that they express is the same. However, getting clear on the nature of propositions is an extremely difficult philosophical task. In this book, I talk about declarative sentences rather than propositions simply because I consider the notion of proposition to be less clear than the notion of a declarative sentence. It is also the case that for the purpose of constructing our formalism, the distinction between declarative sentence and proposition is irrelevant.

11.2 Connecting Sentences: Introducing the Logical Connectives

In English, we use words and phrases like: "and," "or," "not," "therefore," "if. . .then," "hence," "but," "if and only if," and others to connect the components of sentences or to mark a transition from one idea to the next. In sentential logic, we will define logical connectives that look similar in some ways to those ordinary English connecting words. Again, the definitions of the logical connectives in the formalism that we are building will differ from the ordinary natural language usage. The connectives will be defined strictly and explicitly (below), and their role will be to connect declarative sentences to form compound sentences. In natural language, the logical connective words are part of what gives arguments their form and allows us to make inferences. In the formal language of sentential, logic connectives are responsible for *all* of the formal structure of argument.

In this chapter, we will develop a formal system that will allow us to understand how we can legitimately break compound sentences up and join sentences together. For reasons that will be explained in detail below, the logical connectives will be treated as functions. We will assign each of the connectives a symbol and will define them just as we would define a function in mathematics.

11.2.1 And ($\wedge$)

Let's begin with the *and* function. Below, we will introduce symbols for all the relevant logical functions. In the case of the logical function *and*, we will use an upside down wedge symbol "$\wedge$" instead of an English word.

The function which we will define and which we will represent using $\wedge$ differs in some important respects from the English word "and." There are enough basic similarities between the properties of the logical function $\wedge$ and the semantics of the English word "and" that logicians rarely distinguish them. However, as we shall see, the behavior of the conjunction function is somewhat different from the way that the English word "and" behaves in ordinary language.

Let's first begin with the similarities. In ordinary reasoning, if I know that:

1. Lassie is a dog *and* the moon is in Belgrade.

I can conclude that

2. Lassie is a dog.

is true, and

3. The moon is in Belgrade.

is true.

(2) and (3) can be seen to follow from (1) in a way that seems completely obvious and uncontroversial. Given any *conjunction* (a sentence

built out of two sentences joined by the conjunction function), we can validly conclude either side of the conjunction. Even though this is intuitively obvious and seems so fundamental to common sense that it hardly needs an explanation, prior to taking a logic course most of us could not *prove* why (2) and (3) follows from (1). Soon we will be able to prove that we can validly conclude (2) or (3) from (1). Before we are in a position to construct proofs, we will need to formalize our reasoning a bit more.

Using the sentence variables, we introduced above and the logical symbol $\wedge$, we can say that from

$p \wedge q$

I can validly infer

p

I can also validly infer

q

As we saw above, common sense recognizes that patterns of inference like this are legitimate. In fact, this pattern follows what is called ***a rule of inference***. The intuitive rule of inference is sometimes called **the rule of simplification**, and it will be reintroduced later along with a range of other legitimate kinds of moves that we allow in arguments. We will see how one can generate proofs using rules of inferences in later chapters. However, at this early stage, we will be interested in justifying patterns of inference like this. We will begin thinking about how following a rule of inference like simplification can be justified.

Before we move on to other functions, let's explore "and" a little more. The first thing we should notice that conjunctions are false when any of the two or more statements joined by the "and" (the conjuncts) is false. Thus, the compound sentence

Manitoba is in Iowa and Sonora is south of Arizona.

is false, even though it happens to be true that Sonora is south of Arizona. The *conjunction as a whole* is false because of the way the conjunction function works. There are a range of possible roles for words like "and" in ordinary language, some of these roles have nothing to do with the logical function "and." As we have already seen, in building our formalism we must abstract, to some extent, from the ordinary use of language. We restrict ourselves to a specific function rather than attempting to capture the many subtle and interesting things that the word "and" actually does in our language. For example, if I say that

"I woke up and I ate breakfast"

is true, I can not infer that

"I ate breakfast and I woke up"

is true. However, given the way that we will define the logical function $\wedge$ the order of two true statements makes no difference to their resulting truth-value. Therefore, the English word "and" is clearly richer and plays a wider variety of roles than the conjunction function $\wedge$.

In truth functional logic, we will define the behavior of the functions precisely and in so doing we will, of necessity, lose some of the nuance that we find in ordinary language. What we lose in nuance, we gain in clarity.

We will use a symbol define the logical function *and* in terms of truth values in the following way:

Let "and" be represented by "^"
This means that
"*p* ^ *q*" should be read as "*p* and *q*"

We will define "^" as a function which connects two sentence-variables in its own distinctive way.
In order to define the function, we will need to introduce truth tables.

11.3 Truth Tables as Ways of Defining Logical Functions

As we have seen, sentence variables stand for declarative sentences. Declarative sentences are either true or false. As we shall see, when we join two sentence variables together with an $\wedge$, there are four possible ways that they can be true or false together. Before jumping straight into the symbolism, let's analyze a sentence of natural language in order to understand the possible combinations of truth and falsity that we find in conjunctions. Consider the following claim:

> *Saturn's core contains liquid helium* ***and*** *the moon Europa has seas of liquid water.*

This is a compound sentence consisting of two declarative sentences connected by the logical connective "and." As I write this sentence, I confess that I am uncertain whether either of these two declarative sentences is true. Without running a quick Internet search, I have a vague idea that Europa might have seas of liquid water, but I am not sure. I have no idea whether there is helium in Saturn's core. What I know with certainty is that in order for the compound sentence as a whole to be true, both parts must be true.

Without taking any steps to inquire further, I know that:

> "Saturn's core contains liquid helium" is either true or false
>
> "The moon Europa has seas of liquid water" is either true or false

Given two possible truth-values each, we can see that there are four possible ways for two declarative sentences to be true or false together. Here are the four possible combinations:

1. "Saturn's core contains liquid helium" is true and "The moon Europa has seas of liquid water" is true
2. "Saturn's core contains liquid helium" is true and "The moon Europa has seas of liquid water" is false
3. "Saturn's core contains liquid helium" is false and "The moon Europa has seas of liquid water" is true
4. "Saturn's core contains liquid helium" is false and "The moon Europa has seas of liquid water" is false

Therefore, slightly more formally, we can see that while a sentence "*p*" has two possible ways of being true or false, the combination of two sentences "*p*" and "*q*" has four possible ways of being true or false. The philosopher Ludwig Wittgenstein proposed a way of representing the possible combinations of truth-values for compound sentences using what he called schemata (1921). Today, we call Wittgenstein's schemata *truth tables*.

To begin with, let's consider a truth tables consisting of a set of columns falling under sentence variables. The truth table for a single sentence variable *p* consists of a single column falling under the variable in which the two possible truth-values T (for true) and F (for false) are listed:

p
T
F

There are more than two ways that two statements; *p* and *q* can be true and false together. When we construct a truth table to represent all such combinations, we have two columns, one falling under *p* the other falling under *q*. Since there are four possible combinations of truth and falsity for two sentence variables, our table has four rows as follows:

p	*q*
T	T
T	F
F	T
F	F

Row #1 Where *p* is true and *q* is true $(p \cdot q)$
Row #2 Where *p* is true and *q* is false $(p \cdot \neg q)$
Row #3 Where *p* is false and *q* is true $(\neg p \cdot q)$
Row #4 Where *p* is false and *q* is false $(\neg p \cdot \neg q)$

If we had three sentence variables together, our table would be composed of eight rows as follows:

p	*q*	*r*
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

Take a look at these tables and try to determine whether there are any missing combinations of truth-values. You might already be seeing that if there are two truth-values T or F, then the number of possible combinations is 2^n, where n is the number of statements that we are combining. Therefore, for a single statement, there are two possible configurations of truth and falsity, for a pair of statements, there are four; for 3, there are 8; and for 4, there are 16; etc.

You can probably also already see the pattern, by which we organize the possible combinations of true and false cases. Say, we had to list all the possible combinations of truth and falsity for 4 sentence variables: *p*, *q*, *r*, and *s*

To begin with, we know that since there are 4 variables, that there will be $2^4 = 16$ possible configurations (remember that there are 2^n configurations where n is the number of sentence variables).

In order to organize that list of 16 possible configurations, we adopt the following strategy:

> Falling under *p* would be a column consisting of 8 "Ts" followed by 8 "Fs"
> Falling under *q* would be a column consisting of 4 "Ts" followed by 4 "Fs," "followed by 4 "Ts" followed by 4 "Fs"
> Falling under *r* would be a column consisting of 2 "Ts" followed by 2 "Fs," "followed by 2 "Ts" followed by 2 "Fs," "followed by 2 "Ts" followed by 2 "Fs," "followed by 2 "Ts" followed by 2 "Fs"
> Falling under *s* would be a column consisting of alternating "Ts" and "Fs"

You can see why this strategy exhausts all the possible configurations without repeating any configurations. The table of possible combinations for four sentence variables looks like this:

pqrs
TTTT
TTTF
TTFT
TTFF
TFTT
TFTF
TFFT
TFFF
FTTT
FTTF
FTFT
FTFF
FFTT
FFTF
FFFT
FFFF

pqrst
TTTTT
TTTTF
TTTFT
TTTFF
TTFTT
TTFTF
TTFFT
TTFFF
TFTTT
TFTTF
TFTFT
TFTFF
TFFTT
TFFTF
TFFFT
TFFFF
FTTTT
FTTTF
FTTFT
FTTFF
FTFTT
FTFTF
FTFFT
FTFFF
FFTTT
FFTTF
FFTFT
FFTFF
FFFTT
FFFTF
FFFFT
FFFFF

It is important to note that the particular arrangement of Ts and Fs in the columns is simply a conventional matter. Nevertheless, it is a good idea to memorize it at this point. At the very least, you should memorize the arrangement for two variables *p* and *q*.

p	*q*
T	T
T	F
F	T
F	F

The strategy outlined above makes it easy for us to make certain that we have listed all of the possible combinations of truth and falsity for the compound sentence under consideration. At this point, you can predict that if there were five variables *p*, *q*, *r*, *s*, and *t*, there would be 2^5 or 32 rows of combinations. The top row would consist of all case where all the variables are evaluated "true." The last row would be all the cases where the variables were evaluated "false."

Under the first variable—the *p*—there would be 16 Ts followed by 16 Fs.

Under the *q*, we would find 8 Ts, followed by 8 Fs, followed by 8 Ts, followed by 8 Fs.

Under the *r*, we would find 4 Ts, followed by 4 Fs, followed by 4 Ts, followed by 4 Fs, followed 4 Ts, followed by 4 Fs, followed by 4 Ts, followed by 4 Fs.

Under the *s*, we would find 2 Ts, followed by 2 Fs, followed by 2 Ts, followed by 2 F's, followed by 2 Ts, followed by 2 Fs, followed by 2 Ts, followed by 2 Fs, followed by 2 Ts, followed by 2 Fs, followed by 2 Ts, followed by 2 Fs, followed by 2 Ts, followed by 2 Fs, followed by 2 Ts, followed by 2 Fs.

Under the *t*, we would find an alternating column of 32 rows with a single T, followed by a single F, followed by a single Ts, etc.

Returning to the logical function *and*, it will now be possible for us to define it in terms of the four possible combinations of truth-values for two sentence variables. Consider in the next truth table how *and* might work for each of the four combinations of truth and falsity. What would the possible values of a compound sentence be were that compound sentence composed of two sentences joined by the *and* function?

p	$\wedge$	q
T	?	T
T	?	F
F	?	T
F	?	F

In this column are listed the values for *q*.

In this column, we will list the values for $p \wedge q$ as a whole once we figure them out.

In this column are listed the values for *p*.

Our definition of *and* replaces each of the question marks in the table above with either a T or an F. In so doing, we give an exhaustive definition of the behavior of *and* under all possible conditions. The definition of $\wedge$ is pretty straightforward. A compound sentence consisting of two sentences joined by the $\wedge$ function is only true when both parts of the compound sentence are true. It is false in all other cases.

If p is true and q is true, then the sentence as a whole is true.
If p is true and q is false, then the sentence as a whole is false.
If p is false and q is true, then the sentence as a whole is false.
If p is false and q is false, then the sentence as a whole is false.

Thus, we have the following truth table definition for $\wedge$:

p	$\wedge$	q
T	T	T
T	F	F
F	F	T
F	F	F

11.3.1 Or

The English word "or" can serve two principal logical functions as it connects two declarative sentences. Usually in logic, we concentrate on only one of these roles, the so-called inclusive or. We use the symbol "$\vee$" to stand for the inclusive or such that "$p \vee q$" reads "p or q." We will define this function using truth tables along the same lines we did with "and" above:

Intuitively, you might already see that "p or q" is true if either p is true or q is true or both are true. It is false if both p and q are false (again notice that there are only four possible combinations of truth-value).

Let's take a look at the truth table representation of *inclusive or*. Just as we saw in the case of "and" above, we begin by asking what the possible values of a compound sentence would be, if that compound sentence consisted of two sentences were joined by the *inclusive or* function?

p	$\vee$	q
T	?	T
T	?	F
F	?	T
F	?	F

As in the case of *and* above, the definition of *or* replaces each of the question marks with either a T or an F.

If p is true and q is true, then the sentence as a whole is true.

If p is true and q is false, then the sentence as a whole is true.

If p is false and q is true, then the sentence as a whole is true.

If p is false and q is false, then the sentence as a whole is false.

Giving the following truth table definition for *inclusive or*:

p	$\vee$	q
T	T	T
T	T	F
F	T	T
F	F	F

So far we have covered the inclusive sense of the English word *or*. If I mean to use the inclusive sense of the term, then when I say that:

"Students can have cake or ice cream with their coffee."

then I am not denying them the option of having cake *and* ice cream. However, you might think that depending on the context I might mean something else. Earlier, I mentioned that there is another sense of "or" which we use in English, namely, the *exclusive or*. This is the sense of "or" which we would intend when we mean something like "you can have cake or ice cream, but you can't have both". Usually in English, the context indicates that we mean the exclusive rather than inclusive sense *or* such that we don't have to add "and you can't have both." Some languages—Latin, for example—have two different words for each sense of *or*. Latin uses the word "vel" for inclusive, and "aut" for exclusive or.

Consider:

> "I can have another drink or I can drive home without breaking the law."

Can I be sure whether the speaker intends the inclusive or exclusive sense of "or" in this case?
While the table for *inclusive or* looks like this:

P	$\vee$	q
T	T	T
T	T	F
F	T	T
F	F	F

The table for *exclusive or* (sometimes represented by the symbol ⊕) looks like this:

P	$\oplus$	q
T	F	T
T	T	F
F	T	T
F	F	F

Notice that ⊕ only differs from inclusive or in the first case, the sentence which is built from compounding two sentence variables with ⊕ is false when both sentence variables happen to be true, and it is false when both sentence variables are false, but it is true in case one is true while the other is false.

So far we have seen three configurations of truth-values for logical functions which take two truth-values as an argument and give one truth-value as an output. There are many more. In fact there are precisely 16 such functions in total (as we shall see below). At this point, we are in a position to understand why we have been treating the connectives as functions. The way that we have defined them, functions like ⊕, ∨, and ∧ can be understood as taking take two inputs and give one output. The inputs and outputs of these functions are truth-values. This is why sentential logic is sometimes called truth functional logic.

You will recall from your studies of mathematics that functions like

$f(x) = 2x$

is read as saying that the function takes any value and doubles it. A function like this takes a single number as its input and gives another number as the output. For example,

$f(2) = 4$
$f(3) = 6$
$f(11) = 22$

and so on. Sometimes functions are described as being like machines or boxes, where you put in one value and another comes out.

Input output

$2 \quad \boxed{} \Rightarrow \quad 4$

A function is sometimes described as mapping the elements of a domain (in this case the set of numbers) onto a codomain or range.

(domain) $\longrightarrow$ (codomain/range)

$f(x) = 2x$
$1 \longrightarrow 2$
$2 \longrightarrow 4$
$3 \longrightarrow 6$
$4 \longrightarrow 8$
etc.

Some mathematical operations, like addition, subtraction, or multiplication are functions that take a pair of numbers as their input and give a single number as their output. Most truth functions do something similar. They take the truth-values of the pair of sentences that they compound and give a single truth-value as an output. Let's think about the behavior of a truth function like *and*.

$\wedge$

$(T,T) \longrightarrow T$
$(T,F) \longrightarrow F$
$(F,T) \longrightarrow F$
$(F,F) \longrightarrow F$

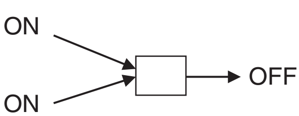
$\wedge$ is a function that takes a pair of truth-values as its input and gives a single truth-value as its output. As we saw above, the truth table definition stipulates the output of all possible combinations of truth-values for the function $\wedge$

This might seem a bit abstract, until you notice that truth functions can be interpreted in some very concrete ways. For example, you can interpret a function like $\oplus$ in the following way:

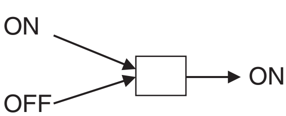
> Imagine a device (we can call it a circuit) which has two incoming wires and one outgoing wire. Now instead of T and F, let's imagine that a wire can be in one of two states; ON or OFF. If a wire is ON, then the wire is carrying an electric current (there are electrons flowing down the wire) and if it's OFF, then it is not carrying a current.

The device that we are about to describe is an xor or $\oplus$ circuit. This is a type of circuit that plays an important role in computation. Here, we see the four possible pairs of inputs and their corresponding outputs for the $\oplus$ circuit. The box represents our truth function and the inputs and output are labeled according to whether or not there is current flowing.

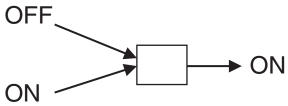
1.



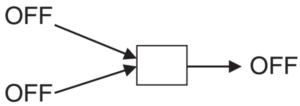
2.



3.



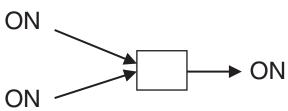
4.



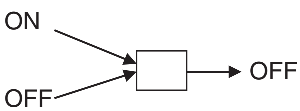
This is the $\oplus$ circuit.

The four drawings above represent the possible combinations of two incoming wires with values of either ON or OFF. The output for each of these configurations will differ depending on which function the circuit is performing. Therefore, for example, the output of the *and* function is ON only in cases where both incoming wires are ON.

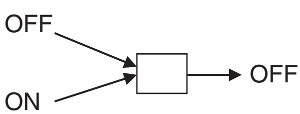
1.



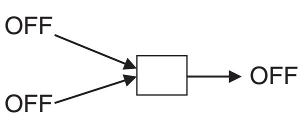
2.



3.



4.



This is the $\wedge$ circuit.

The output of the inclusive or circuit would be ON in all but the final case (4) where both incoming wires are OFF. It should be clear that there are a finite number of possible combinations of circuits of this kind. Similarly, there are a finite number of possible configurations of the binary valued truth table for two variables (4^2). An exhaustive list of all the functions for two variables in sentential logic reminds us of counting from 0 to 15 (the first 16 numbers) in binary. For fun, we can associate the list of binary numbers with the possible configurations of truth and falsity in the following way:

0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
F	F	F	F	F	F	F	F	T	T	T	T	T	T	T	T
F	F	F	F	T	T	T	T	F	F	F	F	T	T	T	T
F	F	T	T	F	F	T	T	F	F	T	T	F	F	T	T
F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T

Each of these columns represents one of the possible 16 functions for two sentence variables. Notice that the 9th combination (1000) is our *and* function (TFFF) and the 15th (1110) is our inclusive *or* (TTTF), the 7th (0110) is our *exclusive or* function (FTTF). From these considerations, it can also be shown that there are only 16 possible binary valued circuits with two incoming wires and one output. Some of the more commonly used logical functions, in addition to *and*, *or*, and, *exclusive or*, are *if. . .then* (1011), *if and only if* (1001), *nand* (0111). These other functions will become familiar to readers in the pages which follow, they are all definable in the same way we defined *and*, *or*, and *exclusive or*.

11.3.2 Not

The word "not" is obviously a critical piece of the logical machinery of ordinary language. When we deny some sentence, we are affirming its negation. We will symbolize the denial of a sentence by placing the symbol "¬" immediately before the sentence. This symbol is sometimes called "negation." Consider the following sentence:

"Xochitl isn't taking a History class."

Clearly, this sentence is denying some state of affairs (that Xochitl is taking a History course) and if we represent

Xochitl is taking a History course.

with the variable

p

and we will represent the denial of the sentence as

$\neg p$

Basically, putting "¬" before a sentence variable has a similar logical function to adding the English phrase "it is not the case that" before another English sentence. Therefore, we can say that "$\neg p$" is true whenever "*p*" is false.

The logical function ¬ is a truth function like the others we have studied except that instead of taking a pair of values it takes the value of a whole sentence and flips it. Take the sentence *p*. It is either true or false. If *p* is true then denying *p* results in a false sentence. If *p* is false then the denial of *p* is a true sentence.

$$(\mathrm{T}) \xrightarrow{\neg} \mathrm{F}$$
$$(\mathrm{F}) \longrightarrow \mathrm{T}$$

Or if we were to represent negation using a truth table it would look like the following:

¬	*p*
F	T
T	F

This is the set of values of *p*

This is the set of values of the negation of *p*

Negation is more interesting than you might imagine. For example, consider a conjunction like "Bill and Hillary were born in Arkansas." This sentence is false. However, if I deny the truth of this sentence what am I committed to? What am I actually saying when I say that it is not true? Well I could be saying that

> Bill was born in Arkansas and Hillary was not born in Arkansas,
>
> or
>
> that Bill was not born in Arkansas and Hillary was born in Arkansas,
>
> or
>
> that Bill was not born in Arkansas and Hillary was not born in Arkansas,

the truth of any of these sentences would make it false that Bill and Hillary were born in Arkansas. Let's consider this example using the truth tables. Let's first remind ourselves of the truth table for the conjunction of two sentence variables. Let *p* be the variable representing "Bill was born in Arkansas" and *q* the variable representing "Hillary was born in Arkansas". The truth table representing the conjunction of *p* and *q* is familiar to us from above:

p	$\wedge$	q
T	T	T
T	F	F
F	F	T
F	F	F

Now, if we were to negate this compound sentence, we would simply add the negation function in a column to the left of the compound sentence.

$\neg$	$(p$	$\wedge$	$q)$
	T	T	T
	T	F	F
	F	F	T
	F	F	F

At this point, we can evaluate the negation of $p \wedge q$ by seeing what happens to the combination of truth-values that are highlighted in green, once we negate them. We end up with the following combination of truth-values which we list under the negation function as follows.

¬	$(p$	$\wedge$	$q)$
F	T	T	T
T	T	F	F
T	F	F	T
T	F	F	F

At this point, we have evaluated the entire compound sentence and we can see that filling in all of the other columns allowed us to evaluate the entire sentence; $\neg (p \wedge q)$. Notice the role played by the parentheses here. We will say more about syntactical function of the parentheses in the next section but for now we should simply notice that putting the negation in front of the parentheses indicates the denial of everything inside the parentheses. Basically, the whole compound sentence "$p \wedge q$" is being denied once we put it in parentheses and add "¬" as its prefix.

Notice also that there are three ways that the denial of $(p \wedge q)$ can be true, the case where p is true and q is false, the case where p is false and q is true, and finally the case where p is false and q is false. The only way that $\neg (p \wedge q)$ can be false is the case where it actually is the case that $(p \wedge q)$, namely the case where p is true and q is true.
The denial of $(p \wedge q)$ is true when:

Row 2 Bill was born in Arkansas and Hillary was not born in Arkansas $(p \wedge \neg q)$

Row 3 Bill was not born in Arkansas and Hillary was born in Arkansas $(\neg p \wedge q)$

Row 4 Bill was not born in Arkansas and Hillary was not born in Arkansas $(\neg p \wedge \neg q)$

Notice that what we have shown is that the negation of the conjunction $(p \wedge q)$ is equivalent to the truth of the disjunction of the second, third, and fourth rows of the truth table. In other words if $(p \wedge q)$ is false then you know that Row 2, Row 3, or Row 4 is true. Being able to read the truth table allows us to see how the same state of affairs can be expressed in a variety of equivalent ways. What we have seen here is that:

$\neg (p \wedge q)$ is equivalent to the assertion of $(p \wedge \neg q)$ or $(\neg p \wedge q)$ or $(\neg p \wedge \neg q)$

In the pages that follow, we will explore the idea of proving the equivalence of declarative sentences using truth tables in more detail.

Negation will play a central role later when we begin to study the logical behavior of the words "all" and "some." By the way of a preview, for example, we will see that the negation of

(A) "It sometimes rains in El Paso."

isn't

(B) "It sometimes doesn't rain in El Paso."

In fact, (A) and (B) are perfectly consistent statements
Instead the negation of (A) is

(C) "It never rains in El Paso."

11.3.3 Implication

In ordinary discussion, we use phrases that tell us about connections between two states of affairs or events. We often hope to express a causal connection, a logical, or a conceptual connection that takes us from one state of affairs or event to another. Consider the following examples:

If you park there, you'll get a ticket.

If I'm confused, then you're really confused.

If she calls back, then you can let her know the news.

If those pants are on sale, get them.

If he had run a marathon, he wouldn't have had that stroke.

The *if. . .then. . .* structure of sentences like these are a pervasive part of our language and our reasoning. Logicians focus on one aspect of this pattern, known as logical implication. The symbol for implication in this text is the horseshoe; "⊃" and from here on we will let the

"if. . .then. . ."

pattern be represented by

"⊃"

such that

"$p \supset q$"

should be read as "if *p* then *q*".

A sentence which is built by connecting two sentence variables with the horseshoe is called a conditional. For example, the sentence

If Jessica is in El Paso, then Sam is in Warsaw.

is a conditional where "Jessica is in El Paso" is the antecedent and "Sam is in Warsaw" is the consequent. While linguistic forms of this kind figure prominently in reasoning, the way that conditionals are understood by logicians is quite different from the usual ways that we think about the conditional in ordinary language and reasoning. Some of these differences result from the way that implication is defined via the truth table. The truth table definition of "⊃" as we saw above corresponds to 1011 or TFTT in the table above.

In tabular form, we define "⊃" as:

p	⊃	*q*
T	T	T
T	F	F
F	T	T
F	T	F

While "⊃" is meant, at least in part, to capture the ordinary sense of linguistic forms like *if p then q* or *p* implies *q*, many philosophers have regarded it as a poor representation of our use of the phrase "if. . .then" in ordinary English. Why? Well, take a look at the truth table for "⊃." It says that any conditional *if p then q* is true except in cases where *p* is true and q is false. This would mean that a conditional which has a true consequent must be true, no matter what the antecedent. But that would mean that the following is true:

> If Martians eat square circles. then this sentence was written after 2011.

Most of us read a sentence like this and think that there is something not quite right about it. The antecedent is false, the consequent is true and in spite of what the truth table definition of "⊃" tells us, a sentence like this cannot be true. We think this because the sense of implication which we would ordinarily understand to be conveyed by sentences like this just doesn't correspond to anything in the real world. Worse still, the following sentence is also true according to the truth table definition of "⊃".

> If Martians eat square circles, then this sentence was written before 2011.

Since it is false that Martians eat square circles (impossible objects are indigestible) and since a conditional with a false antecedent is always true, this sentence is true. Before we despair, it is worth considering the other possible logical functions to determine whether there is another of the 16 possible binary functions which might do a better job capturing our commonsense understanding of the way conditionals work. Once you have carefully considered our other 15 options, we can despair.

In order to grasp the commonsense meaning of "if. . .then" statements, C.I. Lewis introduced the notion of strict implication in 1918. Strict implication is an attempt to include some strong/real necessity into the implication statement. Cases of the kind we considered above, seem to reveal that there is something empty, or purely formal about the way that "⊃" operates in logic. For example, according to the definition of material implication, the only time the conditional is false is in the second case on the truth table where the antecedent is true and the consequent is false. All you're really committed to when you assert the truth of some conditional $p \supset q$ is the denial of the second case in the truth table, namely the relatively lenient claim that

$\neg(p \wedge \neg q)$.

If you're simply saying that the second case doesn't hold, then you're admitting that either

the first case

where p is true and q is true is true,

or the third case

where p is false and q is true is true,

or the fourth case

where p is false and q is false is true.

Logic has sometimes been seen as having only a kind of purely formal necessity; whatever necessity we find in logic is merely a matter of the way we happen to organize our formal systems.

Real necessity is thought to be a property of the real world, not just a property of our formalism. Lewis thought that this real necessity has an important role in the meaning of ordinary "if. . .then" statements. He introduced an additional symbol for possibility, "◊" as well as a special symbol for strict implication "=>" in order to express the difference between material and strict implication.

Lewis hoped to indicate an additional strong level of necessity to his definition such that

$p => q$ (p strictly implies q)

should be read

$\neg\Diamond (p \wedge \neg q)$ (it is not possible that p and $\neg q$)

Lewis's attempt to get clear on the nature of logical implication served as the basis for the development of an entirely new field of logic known as modal logic which studies the concepts of possibility and necessity. By the end of the 20th century, modal logic had given rise to a series of insights concerning nature of necessity that encouraged philosophers to return in a systematic way to the study of metaphysical questions. It is quite striking that reflecting on a relatively simple problem with the logic of the conditional has had such profound consequences for philosophers.

11.4 Building the Language of Sentential Logic

11.4.1 Sentence variables

As we have seen, as a matter of convention and convenience, we let sentences be represented by the lower case letters from the end of the alphabet. In our presentation of sentential logic, "p," "q," and "r" will function as representatives or variable for declarative sentences with definite truth-values. We call these *sentence variables*.

11.4.2 Parentheses

We will use parentheses "(" and ")" to help us keep the syntax of our language unambiguous.

11.4.3 Logical functions

We will assign symbols for logical functions like "and," "not," "if. . .then," "or," etc.

> Let "and" be represented by "$\wedge$." Thus,"$p \wedge q$" should be read as "p and q"
>
> Let "if. . .then. . ." be represented by "$\supset$". Thus, "$p \supset q$" should be read as "if p then q"
>
> Let "or" be represented by "$\vee$." Thus, "$p \vee q$" should be read as "p or q"
>
> Let "not" be represented by "$\neg$." Thus," $\neg p$" should be read as "not p"
>
> Let "if and only if" be represented by "$\equiv$." Thus, "$p \equiv q$" should be read as "p if and only if q"

It is necessary to define these functions precisely. We do this using the truth tables as we have seen above.

11.4.4 Well-Formed formulas (wffs)

The order in which logical functions play their roles is often critical to the meaning of the sentence in which they appear. As in any language, syntactical or grammatical markers are necessary. The parentheses help to mark the syntax of our formulas. Some strings of symbols will be syntactically correct, others will be incorrect. We will only deal with well-formed formulas or wffs. Rules for wffs are as follows:

wff rule #1:
A sentence letter standing by itself counts as a wff. For example, p by itself is a wff

wff rule #2:
The negation of any wff is also a wff. For example, since p is a wff, so $\neg p$

wff rule #3:
When a wff is enclosed by left and right parentheses, the resulting string is also a wff. For example, given some wff p, (p) is also a wff.

wff rule #4:
Given any two wffs, they can be joined by the logical functions $\wedge$, $\supset$, $\vee$, and $\equiv$. For example, if p is a wff and q is a wff, then $p \cdot q$ is a wff. Likewise for $p \supset q$, $p \vee q$, etc.

wff rule #5:
Strings are only wffs if they can be built up from sentence symbols obeying wff rules 1–4 above.

Notice that these rules not only allow us to create strings which are wffs, but they also allow a way of proving whether strings are wffs. For example, if we wanted to prove that some string is a wff of our logical language, we could refer to our rules for wffs in order to determine whether the parts of the string were introduced in accordance with those rules.

11.4.5 Proving whether a string is a wff?

Consider some string:

$(p \supset q) \wedge q$

Is it a wff?

Let's notice that the string consists of the logical function "$\wedge$" connecting two parts of the string (two conjuncts) $(p \supset q)$ and q. It would be a legitimate application of wff rule #4 provided that both conjuncts are wffs. Thus, in order to determine whether this is the case, we have to check to see whether they were built according to wff rules. Since q is a sentence letter standing by itself, it's a wff in accordance with wff rule #1. $(p \supset q)$ is a string which begins with left parenthesis and closes with a right parenthesis. This would be a legitimate application of the wff rule #3 provided that everything inside the parenthesis, namely $p \supset q$ is a wff. $p \supset q$ is a wff according to wff rule #4, just in case p and q are wffs. Sentence letters, by themselves are wffs according to wff rule #1. Therefore, yes, "$(p \supset q) \wedge q$" is a wff according to our rules.

11.4.6 The main logical function (MLF)

In written English, commas, periods, colons, apostrophes, and the like play the role of preventing ambiguity. As we saw in the previous section, in logic, we use parentheses to represent the priority of logical operations. Every compound symbolic sentence in our logic must have an unambiguous name logical function (MLF). A compound sentence, as we've seen is one which contains at least one logical function. For example, while

p
q
r
s

are simple sentences,

$\neg p$
$p \wedge q$
$(p \wedge q) \vee q$

are compound sentences with an unambiguous MLF. Well-formed formulas will have an unambiguous MLF. One of the characteristics of a non-well-formed formula is ambiguity with respect to the interpretation of the roles of logical functions in that formula. By contrast, in a well-formed formula, it will be clear how to evaluate the truth tables for the compound sentence as a whole. In the following section, we will see how to evaluate an entire compound sentence using the truth table method.

11.5 Evaluating Compound Sentences

Take some compound sentence like the following: $(p \equiv q) \supset ((p \supset q) \wedge (q \supset p))$.

It is a well-formed formula with parentheses marking the parts of the formula and their relationships to one another. For example, the part of the formula to the left of the first horseshoe says that p if and only if q

$$(p \equiv q) \supset ((p \supset q) \wedge (q \supset p))$$

To the right of the first horseshoe follows another compound sentence contained by a pair of parentheses on the left and right. The sentence as a whole is a conditional with one compound sentence; $(p \equiv q)$ implying another; $((p \supset q) \wedge (q \supset p))$.

Here, we see how the parentheses form groups of compound sentences. Notice that there are double parentheses immediately following the implication function. In addition, notice that parentheses work as pairs such that whenever a parenthesis is introduced in a well-formed formula, it does so in collaboration with its partner. The two parentheses mark the beginning and end of a well-formed formula which may itself be part of a larger well-formed formula. In ordinary language, it is sometimes impossible to eliminate ambiguity. For example, a sentence like the following allows a range of possible interpretations:

> I have cake on the table and I'll turn on the music if the guests are ready.

The ambiguity that arises in cases like this has to do with our inability to tell which parts of the sentence are grouped with which. For example, is the cake's being on the table dependent on whether the guests are ready, or is the sentence asserting that if the guests are ready then I'll turn on the music, and that, by the way, there is cake on the table independently of whether or not the guests are ready. Well-formed formulas and logic avoid ambiguity of this kind by attempting to be as clear as possible about the ways that the parts of the sentence relate to one another. Returning to our example above, the conditional

$$(p \equiv q) \supset ((p \supset q) \wedge (q \supset p))$$

clearly has as its MLF, the horseshoe connecting the compound sentence to its left with the compound sentence to its right. When we examine each of these compound sentences, we find that they also have MLFs. The MLF of

$$(p \equiv q)$$

is the biconditional "≡," while the MLF of

$$((p \supset q) \wedge (q \supset p))$$

is the *and* function "$\wedge$" joining to compound sentences "$(p \supset q)$" and "$(q \supset p)$." Both of these final to compound sentences have the horseshoe is the MLFs. This analysis provides a guide for how we should evaluate a compound sentence as a whole using the truth tables. The strategy for doing so is straightforward. We begin with the simplest constituent parts of a compound sentence (in our case, these will be the sentence variables) evaluating those before figuring out how the logical functions determine the value of the compound sentence.

Setting up the truth table for a compound sentence is pretty straightforward. The parentheses have no truth-value so they don't need their own columns. Everything else will get a column falling under it. In this case, since there are only two sentence variables, we know that we will only have to include four rows in our table. This is because, as we saw above, the number of rows needed in order to exhaust all possible combinations of truth-values for n variables is 2^n

$(p$	$\equiv$	$q)$	$\supset$	$((p$	$\supset$	$q)$	$\wedge$	$(q$	$\supset$	$p))$

Beginning with a blank table, we assign conventional set of truth-values underneath each of the sentence variables. We must be sure to do so consistently so that we end up with the following

$(p$	$\equiv$	$q)$	$\supset$	$((p$	$\supset$	$q)$	$\wedge$	$(q$	$\supset$	$p))$
T		*T*		*T*		*T*		*T*		*T*
T		*F*		*T*		*F*		*F*		*T*
F		*T*		*F*		*T*		*T*		*F*
F		*F*		*F*		*F*		*F*		*F*

At this point, once we've assigned values to the simplest parts of the formula (the sentence variables), we must decide on where to begin determining the truth-value of the various compound sentences that make up the larger compound sentence as a whole.

$(p$	$\equiv$	$q)$	$\supset$	$((p$	$\supset$	$q)$	$\wedge$	$(q$	$\supset$	$p))$
T		*T*		*T*		*T*		*T*		*T*
T		*F*		*T*		*F*		*F*		*T*
F		*T*		*F*		*T*		*T*		*F*
F		*F*		*F*		*F*		*F*		*F*

We've already seen that this horseshoe is the MLF for the entire sentence. Once we have the set of values falling under the MLF, we have the value of the sentence as a whole. We recall our earlier analysis of the sentence where we found that the relative priority of functions could be pictured along the following lines

$$(p \equiv q) \supset ((p \supset q) \wedge (q \supset p))$$

The MLF in the formula above is the horseshoe. It connects the following two formulas:

"$p \equiv q$"

and

"$(p \supset q) \wedge (q \supset p)$"

Those two formulas have their own MLF, in the case of "$p \equiv q$" the MLF is the "≡" joining p with q

In the case of $(p \supset q) \wedge (q \supset p)$, the MLF is the "∧" joining "$(p \supset q)$" with "$(q \supset p)$"

Likewise, "$(p \supset q)$" and "$(q \supset p)$" each have their own MLF. In both cases, the MLF is a horseshoe

Here, we can see the relative ordering or priority of functions. Our use of the truth table begins with most basic or lowest priority functions and builds up to the point where we can evaluate the MLF of the compound sentence. The functions connecting individual sentence variables are the first that we evaluate. In our case, reading from the left, the biconditional and the two horseshoes are highlighted in pink. As we begin filling in the values in our truth table, we refer to the definition of the functions in order to know how to complete each place in the column. For example, in the case of the first biconditional, we know from the definition of the biconditional function that it gives T as a value in cases where it takes is an input a pair of Ts or a pair of Fs. It gives F as a value otherwise. Following this definition, we can begin to evaluate the column falling under the biconditional.

$(p$	$\equiv$	$q)$	$\supset$	$((p$	$\supset$	$q)$	$\wedge$	$(q$	$\supset$	$p))$
T	T	T		T		T		T		T
T	F	F		T		F		F		T
F	F	T		F		T		T		F
F	T	F		F		F		F		F

I have drawn the line through the truth-values falling under *p* and *q*, in this case to make it clear that the resulting truth table falling under the biconditional symbol is the only one which represents the truth-value of ($p \equiv q$). From there, let's fill in the values for the first horseshoe as follows

$(p$	$\equiv$	$q)$	$\supset$	$((p$	$\supset$	$q)$	$\wedge$	$(q$	$\supset$	$p))$
T	T	T		T	T	T		T		T
T	F	F		T	F	F		F		T
F	F	T		F	T	T		T		F
F	T	F		F	T	F		F		F

Filling in the values for the second horseshoe provides us with an opportunity to see how we can apply the definition of the function in cases that aren't as obvious as our first two.

$(p$	$\equiv$	$q)$	$\supset$	$((p$	$\supset$	$q)$	$\wedge$	$(q$	$\supset$	$p))$
T	T	T		T	T	T		T	T	T
T	F	F		T	F	F		F	T	T
F	F	T		F	T	T		T	F	F
F	T	F		F	T	F		F	T	F

Notice how, in the third row falling under the second horseshoe, we have an F rather than a T. This is because the two pairs of values are different in the cases of the first and second horseshoe. The order makes a difference, and even though the same variables and the same function feature in both, "“if *p* then *q*” is true in a different way than “if *q* then *p*.”

At this point, we are just two steps away from evaluating the entire compound sentence. We notice that the *and* function serves to connect the two conditionals on the right-hand side of the MLF. In order to evaluate the result of joining these two conditionals, we take the pair of values of the conditionals and apply the true function *and*. In the table below, I've drawn arrows showing the two columns which serve as the basis for determining the value falling under *and.*

(*p*	≡	*q*)	⊃	((*p*	⊃	*q*)	∧	(*q*	⊃	*p*))
T	T	T		T	T	T	T	T	T	T
T	F	F		T	F	F	F	F	T	T
F	F	T		F	T	T	F	T	F	F
F	T	F		F	T	F	T	F	T	F

With this step, we calculated the value of the compound sentence to the left of the MLF. Since we already have the value of the compound sentence to the right of the MLF, namely $p \equiv q$, all that remains is to evaluate the conjunction of the values highlighted in green below. This is done as follows, for every row of values take the pair of truth-values and apply the true function. We know that the rule for the horseshoe is to give T as a value in all cases except when the antecedent is true and the consequence is false. Since there are no cases where the first value is T. and the second value is F, this entire complex conditional statement is always true. Thus under the MLF, we see a column of Ts and no Fs.

(*p*	≡	*q*)	⊃	((*p*	⊃	*q*)	∧	(*q*	⊃	*p*))
T	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*
T	*F*	*F*	*T*	*T*	*F*	*F*	*F*	*F*	*T*	*T*
F	*F*	*T*	*T*	*F*	*T*	*T*	*F*	*T*	*F*	*F*
F	*T*	*F*	*T*	*F*	*T*	*F*	*T*	*F*	*T*	*F*

As we have seen, statements which are true under all conditions by virtue of their form are called tautologies. In this case, we can see that any sentence which has this form will always be true. Another way of putting this is to say that everything to the right of the horseshoe follows logically from the sentence to the left of the horseshoe. Still one more way of putting this is to say that the compound sentence to the right of the horseshoe is validly implied by the compound sentence to the left of the horseshoe.

We can now see how truth tables can serve as a way of deciding whether an argument in sentential logic is valid. Given some set of premises and a conclusion, treat the premises as a conjunction. This conjunction is the antecedent of a conditional whose consequent is the conclusion. The entire argument now forms a single formula. Run the truth table for the whole statement. If the value of the statement falling under the MLF is true in all cases, the conditional is a tautology, and the argument is valid.

11.5.1 Some examples

In section 11.2.1, we encountered a simple argument that ran as follows:

If I know that:

1. Lassie is a dog *and* the moon is in Belgrade.

I can conclude that

2. Lassie is a dog

is true, and

3. The moon is in Belgrade

Is true

I mentioned above that this is an inference that follows a rule of inference known as the rule of simplification, and I promised that by the end of the chapter, we would be able to prove that this rule is reliable and legitimate. Let's recall that simplification is a rule of inference for sentences with the "$\wedge$" symbol as the main logical operator. The rule states that whenever a sentence of the form $p \wedge q$ is true, then p is true and q is true. In schematic form, simplification can be presented as (the symbol "$\therefore$" stands in for "therefore"):

$p \wedge q$

$\therefore p$

Now that we have studied the truth tables we are in a position to prove that this rule will not lead us astray under any circumstance.

Let's imagine that someone has followed the rule of inference and has moved from $p \wedge q$ to p. Is his inference valid?
As described above, we take the premises and join them to the conclusion with the horseshoe to form a single sentence

$(p \wedge q) \supset p$

We can now evaluate the sentence as a whole using the truth table method. We begin by setting up our conventional ordering of truth-values for all the variables:

$(p$	$\wedge$	$q)$	$\supset$	p
T		T		T
T		F		T
F		T		F
F		F		F

At this point, we identify the order of the truth functions using the parentheses. The main logical operator is $\supset$ and the secondary logical operator is $\wedge$, thus we evaluate the $\wedge$ first before evaluating $\supset$.

$(p$	$\wedge$	$q)$	$\supset$	p
T	**T**	T		T
T	**F**	F		T
F	**T**	T		F
F	**T**	F		F

Finally, we see that the truth-value of the sentence as a whole, falling under the main logical operator $\supset$ is a tautology.

$(p$	$\wedge$	$q)$	$\supset$	p
T	T	T	**T**	T
T	F	F	**T**	T
F	T	T	**T**	F
F	T	F	**T**	F

What this means is that this pattern of reasoning can never be false under any possible condition. We have thereby demonstrated that sentences of this form are logically true. Let's take another simple example. Consider an argument like the following:

$(p \supset q)$
p
$\therefore q$

This is an instance of a now-familiar pattern of inference; Modus Ponens. How do we know we can rely on it? Again we can use to the truth table method to check. This time we have two premises $(p \supset q)$ and p. The conclusion of the argument is q. We can set up the argument as a sentence whose MLF is the horseshoe as follows. First, list the conventional ordering for the variables

$((p$	$\supset$	$q)$	$\wedge$	$p)$	$\supset$	q
T		T		T		T
T		F		T		F
F		T		F		T
F		F		F		F

Then solve for the first $\supset$

$((p$	$\supset$	$q)$	$\wedge$	$p)$	$\supset$	q
T	T	T		T		T
T	F	F		T		F
F	T	T		F		T
F	T	F		F		F

Then solve for $\wedge$

$((p$	$\supset$	$q)$	$\wedge$	$p)$	$\supset$	q
T	T	T	T	T		T
T	F	F	F	T		F
F	T	T	F	F		T
F	T	F	F	F		F

Finally, solve for the MLF; in this case, the second ⊃

$((p$	$\supset$	$q)$	$\wedge$	$p)$	$\supset$	q
T	T	T	T	T	**T**	T
T	F	F	F	T	**T**	F
F	T	T	F	F	**T**	T
F	T	F	F	F	**T**	F

Thus revealing that the sentence as a whole is a logical truth, or a tautology.

11.5.1.1 The Truth Table Method of Proof

The truth tables provide our first method for proving validity. In cases where the column falling under the horseshoe has only T's, we say that the argument has no counterexamples. This means that the conclusion follows from the premises under all possible combinations of truth-values for the premises. This, of course, means that it is a valid argument. In cases where there the line falling under the horseshoe has F's, those cases represent counterexamples to the argument. This is the case for an invalid argument.

Let's see how we can read the counterexamples off our truth table for an invalid argument. Take as an example of a fallacious piece of reasoning an instance of the fallacy of affirming the consequent.

> If the surface is colored, then it has a shape.
>
> It has a shape.
>
> Therefore, it is colored

We can easily recognize that this is a bad piece of reasoning. We can see that it is an instance of the following form:

$p \supset q$
q
$\therefore p$

At this stage, we have a method for evaluating its validity, namely by constructing a truth table to determine whether the argument (rewritten as a compound sentence) is logically true. The compound sentence, we recall is a conditional composed of the conjunction of the premises as the antecedent and the conclusion as the consequent.

$((p \supset q) \quad q) \supset p$

When we run the truth table for the sentence, we find that the sentence as a whole is not valid; not always true. Instead, it is false in the third row, the scenario in which p is false and q is true. This case constitutes a counterexample.

$((p$	$\supset$	$q)$	$\wedge$	$q)$	$\supset$	p
T	T	T	T	T	T	T
T	F	F	F	F	T	T
F	**T**	**T**	**T**	**T**	**F**	**F**
F	T	F	F	F	T	F

Here is the counterexample. What it tells us is that since sometimes it is false that p and true that q (in the third row of our conventional ordering of truth-values for p and q) it is not the case that given $p \supset q$ and q, p must follow. The third row of the truth table is the counterexample to the argument. The truth table method shows us the scenario in which the compound sentence as whole is false. In other words, to return to our natural language example, we now see why the following argument is invalid.

If the surface is colored then it has a shape

And

It has a shape

Therefore

it is colored

We cannot validly conclude that because surface has a shape and if a surface is colored it has a shape, *therefore* it is colored. Quite simply, there are circumstances, like the third row on the truth table were things have shape and are not colored. The conclusion of a valid argument follows of necessity from its premises. Since this conclusion does not necessarily follow, the argument is invalid.

11.5.2 Normal forms

In 1913, Henry Sheffer proved that all the truth-functional functions that we have studied so far and any combination of these functions in a well-formed formula can be expressed by a single function. The function he identified is now called the Sheffer stroke or NAND and is usually represented with a $|$ or a $\uparrow$ and is equivalent to the denial of the conjunction.

$p \mid q$

is equivalent to

$-(p \wedge q)$

The truth table for the Sheffer stroke is:

p	$\mid$	q
T	F	T
T	T	F
F	T	T
F	T	F

Given any well-formed formula of sentential logic, we can find a logically equivalent formula which consists only of sentence letters and Sheffer strokes. To convince yourself of this, the first step is to think about the way the truth table represents each of the truth-functional functions.

Recall that each row in the truth table can be understood as a conjunction:

Row #1 Where p is true and q is true $(p \wedge q)$
Row #2 Where p is true and q is false $(p \wedge -q)$

Row #3 Where p is false and q is true $(-p \wedge q)$
Row #4 Where p is false and q is false $(-p \wedge -q)$

Now recall that the definition of a function consists solely in asserting which of the rows is true and which is false for that function. For the case of the material conditional, the horseshoe, we say that it is true in the first third and fourth row, but false in the second.

Row #1 T
Row #2 F
Row #3 T
Row #4 T

Recall that we defined all 16 functions similarly and that the horseshoe was the 12th of our functions.

0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
F	F	F	F	F	F	F	F	T	T	T	T	T	T	T	T
F	F	F	F	T	T	T	T	F	F	F	F	T	T	T	T
F	F	T	T	F	F	T	T	F	F	T	T	F	F	T	T
F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T

At this point, recall that we can simply read off the disjunctive normal form from the truth tables. For example, material conditional, or the horseshoe, can be thought of as a truth table in which only the second row contains an F. We know that "$p \supset q$" can be represented by

$(p \wedge q) \vee (-p \wedge q) \vee (-p \wedge -q)$ [true in the first or third or fourth row]

Alternatively, we might notice that it is false only in the second case. Denying the second row means asserting:

$-(p \wedge -q)$

At this point, we can see that the truth table allows us to immediately convert any formulas in sentential logic into formulas containing negation, conjunction, and disjunction. Since we can express disjunctions with conjunctions and negations, we see that a normal form consisting of nand (the negation of a conjunction) will suffice to represent all logical functions.

Showing how a disjunction can be represented in terms of nand

$(p \vee q)$
is equivalent to
$\neg\neg(p \vee q)$ [By a rule of inference known as double negation]
which is equivalent to
$\neg (\neg p \wedge \neg q)$ [By a rule of inference known as DeMorgan's law]
which is equivalent to
$\neg (\neg(p \wedge p) \wedge \neg(q \wedge q))$ [Since any wff is equivalent to a conjunction of that wff with itself]
$\neg(p \wedge p) \mid \neg(q \wedge q)$ [Replacing the negation and conjunction with a nand]
$(p \mid p) \mid (q \mid q)$ [Replacing the remaining negations and conjunctions with a nand]

11.6 Effective Procedures and Decidability

In this chapter, we learned a formal method; the truth table method that provides a reliable test for validity in cases where declarative sentences are connected by simple logical operators. If we are asked whether a particular argument in sentential logic is valid, the truth table method gives us a recipe for deciding. More specifically, we can know that there is a definite "yes" or "no" answer that we will discover in a finite number of steps by following the recipe.

Imagine a complete novice to human kitchens, an alien perhaps, who you must instruct. The task is to prepare breakfast. The alien is not very intelligent but can understand some basic instructions. A recipe written for someone who has no idea what food is, what the ingredients are, how to use a pot, stove, etc. would be a challenge. Fortunately, our task in this chapter has been simpler. We have developed a mechanical procedure; a completely explicit recipe, for deciding whether a sentence of sentential logic belongs to the set of logically true sentences. Philosophers and logicians call recipes of this kind ***effective procedures***. An effective procedure has the following characteristics:

- It finishes after a finite number of steps.
- It always produces the correct answer.
- It consists of a finite number of instructions each of which is explicit and can be completed in a finite amount of time.

If there is an effective procedure that can solve all instances of a well-defined problem, philosophers and logicians call that problem **decidable**. The truth table method shows that the question of whether an argument of sentential logic is valid is decidable. Decidability is an important feature of sentential logic. It is an impressive fact that there is an **effective procedure** for settling the question of validity for sentential logic in a finite number of steps.

12 Deductive Proof in Sentential Logic

Now that we have the basics of our formalism in place, let's return to our informal definition of arguments from earlier. As we saw previously, an argument is a sequence of sentences that is meant to convince an audience of some claim or conclusion. It is not always easy to identify the premises (or sometimes the conclusion) of arguments, but in a good argument, we say that the conclusion *follows from* true premises. We saw that an argument is valid, when it is not possible for the premises of the argument to be true and its conclusion to be false. Unaided common sense, as we have seen in previous chapters is not a reliable guide to determining the validity of sentences.

Truth tables are great. In fact, it is possible to prove that truth tables provide the basis for an effective procedure for determining the validity of any argument in sentential logic. However, truth tables are not always the easiest or practical way to actually prove things. For example, laying out a truth table by hand gets a bit cumbersome once an argument contains, for example, more than four declarative sentences. Avoiding errors is challenging with 16 or more rows on a truth table. It is not unusual for ordinary arguments to have this level of complexity. As we saw in the previous chapter, given sufficient patience and care, we can use the truth tables to allow us to determine whether any argument in sentential logic is valid.

Happily, there are other ways to efficiently decide whether an argument in sentential logic is valid, which do not involve the laborious task of listing all the possible combinations of truth and falsity on a truth table. In this chapter, we will examine additional methods for proving validity. We will construct **direct and indirect proofs** using **rules of inference** and we will learn a mechanical technique called the **tree method**. By the end of this chapter, we will see that all these methods of proof are connected.

12.1 Following from

If an argument is valid we sometimes say that its conclusion *follows from* or is a *logical consequence* of its premises. The idea of one sentence *following from* another is somewhat vague, but it can be made precise along the following lines:

> Conclusion B *follows from* the set of premises A
>
> if one can legitimately move from A to B using steps, which abide by the rules of inference
>
> if the moves which get you from A to B are all logically justifiable

This idea of a proof as a derivation or chain that leads from the premises to the conclusion might be somewhat familiar to you from your study of geometry. In high school geometry, theorems are proven step-by-step from axioms. We say that the theorem, say Pythagoras' theorem can be derived, step-by-step, from the axioms and some rules that we regard as holding for geometric reasoning. We will use a similar approach to proving the validity of arguments in sentential logic. We will construct derivations of conclusions from premises using only truth preserving rules of inference. These derivations will be:

- A sequence of sentences.
- Each sentence is a premise or else has been deduced from one or more of its preceding members by means of the application of a legitimate rule of inference.
- The last member of the sequence will be the conclusion of the argument.

If we can construct a derivation using only legitimate rules of inference, then we can know for sure that the argument is valid.

Another method that we will examine in this chapter is called the **tree method**. It will be obvious that the tree method is closely related to the truth table method that we saw in Chapter 11. The tree method relies upon some additional assumptions that we will discuss in detail below. The most important of these involves so-called **reductio** arguments. This method is appears quite different from the kinds of deductive proofs we will examine initially. However, both methods shed light on aspects of the others.

12.2 Rules of Inference

As we introduce the rules of inference, it's worth noting briefly that each rule of inference is simply a valid pattern of inference. In sentential logic, we can treat each pattern as a schema. What we mean by schema is something like a structure, outline, or framework. The great American philosopher W. V. Quine called schemata "logical diagrams of sentences" (1982, 33).[1] He highlighted the role of variables in this context, pointing out that "the letters '*p*', '*q*', etc., by supplanting the component clauses of a statement serve to blot out all the internal matter which is not germane to the broad outward structures with which our logical study is concerned." (Ibid, 33).

In this section, we will introduce nine rules of inference. There are an infinite number of valid patterns of inference. This sample represents those that resemble the patterns we frequently find in ordinary reasoning. It is also the case that these rules are the ones that are traditionally mentioned by logicians.

12.2.1 Simplification

Let's consider, for example, our first rule of inference, the rule of simplification. As we have already seen, this rule can be stated as follows:

> *If we are given the conjunction of two wffs than we can infer both of those wffs in isolation.*

The sense in which it is a schema can be understood as follows: As we construct derivations, we can call on the rule of simplification

[1] Quine, W. V. O. (1982). *Methods of logic*. Harvard University Press.

whenever we see formulas with the "$\wedge$" symbol as the main logical function. The rule states that whenever a sentence of the form $p \wedge q$ is true, then p is true and q is true. In schematic form, simplification is defined as (the symbol "$\therefore$" stands in for "therefore"):

$p \wedge q$
$\therefore p$

or

$p \wedge q$
$\therefore q$

That is, from a conjunction one may derive either the left-hand conjunct or the right-hand conjunct. Let's use this rule in the context of a derivation. This will be our first example of how to use a rule of inference in a derivation.
Let the sentence constants P and Q be defined as follows:

P = "Quine is a philosopher."
Q = "Quine wrote *Methods of Logic*."

The sentence "Quine is a philosopher and he wrote *Methods of Logic*." can be written in the language of sentential logic as:

$P \wedge Q$

Since $P \wedge Q$ is true, we also know that P is true by applying the inference rule simplification. We can prove P follows by constructing a derivation:

Derivation 1: Show P

1. $P \wedge Q$	Premise
2. P	S, 1

Our derivations will follow the two-column model of proofs in geometry. In the left-hand column, we have a numbered list the moves in our proof. In the right, we explain how we can legitimately assert the corresponding move. Derivation 1 consists of two lines. Line 1 is the premise and line 2 is the result of applying the Rule of Simplification to Line 1. Line 2 is also the conclusion of Derivation 1. Notice that each line is justified by the comment in the right hand column. Line 1 gets to be there because it is a premise of the argument. Line 2 because it is the result of applying a legitimate rule The Rule of Simplification (S) to a preceding line of the proof. We can apply the Rule of Simplification because the "$\wedge$" symbol is the main logical function of line 1.

SIMPLIFICATION (S.)

$p \wedge q$
therefore
p

$p \wedge q$
therefore
q

In Section 11.5.1 we demonstrated using the truth table why we can be quite confident that applying the rule of simplification will never lead us astray. The rule of simplification is a truth preserving pattern of inference. All of the rules of inference that we will introduce in this chapter can be tested in this way and all will be valid.

We can confidently use the rule of simplification in proofs. Whenever we encounter a conjunction, we can, if we wish, derive either side of the conjunct on a line by itself. For example, we can prove that q follows from $(p \wedge q) \wedge (r \supset s)$ by applying the rule of simplification twice. The proof would look like this:

Derivation 2: Show q

1.	$(p \wedge q) \wedge (r \supset s)$	Premise
2.	$(p \wedge q)$	S, 1
3.	q	S, 2

As described above, we number the lines and in the right-hand column, we explain why each line is allowed to appear. The premise appears on line one, the rule of simplification allows me to infer the second line because the $\wedge$ is the main logical function of the sentence. Applying the rule of simplification to the resulting line 2 allows us to infer the third line. The third is the conclusion. These three lines constitute a proof that q follows from $(p \wedge q) \wedge (r \supset s)$.

12.2.2 Conjunction

The Rule of Conjunction is another intuitively obvious rule of inference whose validity can be demonstrated using the truth table method if you feel so inclined. This rule says that any two previous lines in a proof can be joined together with the logical *and* to form a new line. This move is another valid pattern of inference.

CONJUNCTION (CONJ)

p

q

therefore

$p \wedge q$

Let's use the rule of conjunction to prove that we can derive

$(r \wedge m)$

From two premises:

$(p \wedge q) \wedge r) \wedge (q \supset s)$

$(q \supset s) \wedge m$

Before we generate the proof, it is worth locating the main logical functions of both premises (in both cases the MLF is $\wedge$). Then it is worth thinking through what we are attempting to derive as our conclusion, namely $(r \wedge m)$. At this point, our strategy should be to think about how we can build a conjunction like $(r \wedge m)$. Since we only have two rules in our inventory at the moment, Simplification and Conjunction, we recognize that in order to build $(r \wedge m)$ we will need to first have both conjuncts r and m on lines by themselves before applying the rule of conjunction to conclude $(r \wedge m)$. In order to get r and m on lines by themselves we will need to apply rules of inference to the premises. Again, at this point we only have two rules of inference; Simplification and Conjunction, therefore in order to decompose the premises, we will be applying Simplification. Below we will see what this derivation looks like:

Derivation 3: Show $(r \wedge m)$

1.	$((p \wedge q) \wedge r) \wedge (q \supset \mathrm{s})$	Premise
2.	$(q \supset \mathrm{s}) \wedge m$	Premise
3.	m	S, 2
4.	$(p \wedge q) \wedge r$	S, 1
5.	r	S, 4
6.	$r \wedge m$	Conj, 5 & 3

12.2.3 Disjunctive Syllogism

The next rule we will introduce; Disjunctive Syllogism is an inference rule for the "$\vee$" operator. The rule states that if a disjunction is true and one of its disjuncts is false, then the other disjunct must be true. A disjunct is one of the sentence variables or constants of a disjunction connected by the disjunction operator "$\vee$". In schematic form, disjunctive syllogism is defined as:

$$\begin{array}{c} p \vee q \\ \neg p \\ \therefore q \end{array} \qquad \text{OR} \qquad \begin{array}{c} p \vee q \\ \neg q \\ \therefore p \end{array}$$

In other words, from a disjunction and the negation of one of its disjuncts, one may derive the other disjunct. Consider the following argument:

1. Either Quine is a philosopher or he wrote *Word & Object*.
2. Quine is not a philosopher.
3. Therefore, Quine wrote *Word & Object*.

Using the same sentence constants as those used in Derivations 1 and 2, we can construct the following proof:

Derivation 4: Show Q

1. $p \vee q$	Premise
2. $\neg p$	Premise
3. Q	DS, 1 & 2

Unlike conjunctions, which permit one to derive each conjunct, we cannot immediately derive each disjunct from a disjunction. Rather, given a true disjunction we do not know which disjunct is true. One or both disjuncts may be true, but we do not know which.
Supposing that the sentence constants stand in for English sentences, consider the following derivation:

Derivation 5: Show E

1. $((A \wedge B) \vee (C \wedge D))$	Premise
2. $((\neg C \vee E) \wedge \neg(A \wedge B))$	Premise
3. $\neg(A \wedge B)$	S, 2
4. $\neg C \vee E$	S, 2
5. $C \wedge D$	DS, 1 & 3
6. C	S, 5
7. E	DS, 4 & 6

In line 3, using simplification, we were able to derive one of the conjuncts from line 2. However, we cannot immediately derive one of the disjuncts from line 1. Rather, as line 5 shows, we needed line 3 in order derive the disjunct on line 5 using disjunctive syllogism.

DISJUNCTIVE SYLLOGISM (DS)

$p \vee q$

$\neg p$

therefore

q

$p \vee q$

$\neg q$

therefore

p

12.2.4 Addition

This is simply the rule of inference, which says that, using the *inclusive or* one can join any sentence to a sentence which has appeared in a previous line in a proof. For example, if it's true that Everest is the tallest mountain then by using the rule of addition I can conclude that Everest is the tallest mountain or your logic teacher is a Texas sharpshooter. This rule of inference feels a bit strange to most of us, but again, if you feel any serious doubt you can always check the validity of the pattern using a truth table. Addition is our second inference rule involving the "$\vee$" operator. Put simply, if some sentence p is true, then $p \vee q$ is true, where q is any declarative sentence whatsoever. In schematic form, addition is defined as:

p

$\therefore p \vee q$

Thus, one may simply add any (simple or complex) sentence to any previous sentence by connecting it to the previous sentence with the "$\vee$" operator. Consider the following argument:

1. Georg Cantor was a brilliant mathematician.
2. Therefore, Georg Cantor was a brilliant mathematician or he is the king of France.

This may not seem like much of an argument, but it is provably valid using the method we learned in Chapter 11. Moreover, the argument is *sound*, that is, it is valid and has true premises. Letting C stand in for the first sentence in the above argument, and given that C is true, we can correctly infer that C or any sentence you can imagine is true. Thus, consider the following example:

Derivation 6: Show $(C \vee ((A \wedge B) \vee (D \wedge E))$

1.	C	Premise
2.	$(C \vee ((A \wedge B) \vee (D \wedge E))$	Add, 1

One may wonder why we would include a rule like addition into our formal system of logic, so consider the following argument:

1. It is the case that if either Georg Cantor was a brilliant mathematician or he is the king of France, then he invented set theory.
2. Georg Cantor was a brilliant mathematician.
3. Therefore, Georg Cantor invented set theory.

It would be unnecessarily difficult and cumbersome to demonstrate the validity of this argument without a rule like addition. We also need to rules of inference for conditionals as the main logical operator to construct a proof of the above argument. Therefore, we will return to this argument later.

ADDITION (ADD)

p

therefore

$p \vee q$

12.2.5 Modus Ponens

Modus Ponens is an inference rule for the material conditional or the "$\supset$"operator. The rule states that if you have a true conditional statement and its antecedent is true, then its consequent is true. Or, in other words, from a conditional and its antecedent one can derive its consequent. In schematic form, modus ponens is defined as:

$p \supset q$

p

$\therefore q$

In the above schema, the left side of the conditional (p in this case) is called its *antecedent*. The right side of a conditional statement (q in the above case) is called its *consequent*. Since the conditional and its antecedent are taken as premises, the consequent follows straight away. Consider, again, the following argument:

1. It is the case that if either Georg Cantor was a brilliant mathematician or he is the king of France, then he invented set theory.
2. Georg Cantor was a brilliant mathematician.
3. Therefore, Georg Cantor invented set theory.

Let the following sentence constants stand in for the simple sentences above:

C = Georg Cantor was a brilliant mathematician.

F = Georg Cantor is the king of France.

S = Georg Cantor invented set theory.

Then, we can construct the following derivation (note the use of addition):

Derivation 7: Show S

1.	$((C \vee F) \supset S)$	Premise
2.	C	Premise
3.	$C \vee F$	Add, 2
4.	S	MP, 1 & 3

At this point, we can run a proof that employs three of these rules of inference in a proof. Consider the following scenario. You are told three things about a game of chess. These are your premises. From these three pieces of information your task is to determine whether

Either the white pawn will move forward or the white queen will take the knight.

It is not the case that the white pawn will move forward and the black bishop will be sacrificed.

If the white queen takes the knight then if the black bishop is sacrificed, checkmate, black wins.

P = pawn will move forward

Q = queen will take the knight

S = bishop will be sacrificed

M = checkmate, black wins

Derivation 8: Show M

1.	$P \vee Q$	Premise
2.	$-P \wedge S$	Premise
3.	$Q \supset (S \supset M)$	Premise
4.	$-P$	S, 2
5.	Q	DS, 1 & 4
6.	$(S \supset M)$	MP, 3 &5
7.	S	S, 2
8.	M	MP, 6 & 7

MODUS PONENS (MP)

$(p \supset q)$

p

therefore

q

12.2.6 Modus Tollens

Like Modus Ponens, Modus Tollens is a rule of inference for sentences containing the "⊃" symbol as their main logical operator. For any conditional statement, if the conditional is true and its consequent is false, then the antecedent is false. In schematic form, modus tollens is defined as:

$p \supset q$

$\neg q$

$\therefore \neg p$

In other words, from a conditional and the negation of its consequent, one may derive the negation of its antecedent. Consider the following argument:

1. If it's raining, then it's cloudy.
2. It is not cloudy.
3. Therefore, it is not raining.

If we translate the above argument into the language of sentential logic, assigning appropriate sentence constants to stand in for simple sentences, we can construct the following derivation:

Derivation 9: Show $\neg R$

1.	$R \supset C$	Premise
2.	$\neg C$	Premise
3.	$\neg R$	MT, 1 & 2

Since it rains only if it's cloudy and it's not cloudy, then it's not raining. It should be clear that from a conditional alone, one cannot derive either side of the conditional. Modus ponens states that from a conditional and its antecedent, one can derive its consequent, and modus tollens states that from a conditional and the negation of its consequent, one may derive the negation of its antecedent. Therefore, as a matter of strategy, when you see any sentence whose main logical operator is the "$\supset$" symbol, look for its antecedent or the negation of its consequent when constructing a derivation.

12.2.7 Hypothetical Syllogism

Hypothetical Syllogism is another rule of inference for the "$\supset$" operator. From two conditional statements where the consequent of the first conditional is the same as the antecedent of the second, one may derive a third conditional whose antecedent is the antecedent of the first conditional and whose consequent is the consequent of the second conditional. Stating this rule in English is a bit cumbersome and at this stage you are probably becoming pretty comfortable with seeing the rules presented formally. Nevertheless, the rules says that if you are given two conditionals where the consequent of the first conditional is the same as the antecedent of the second conditional, one may derive a third conditional whose antecedent is the left-hand side of the first conditional and its consequent is the right-hand side of the second conditional. At this point, the formal representation is more obvious and elegant than the English:

HYPOTHETICAL SYLLOGISM (HS)

In schematic form, hypothetical syllogism is defined as:

$p \supset q$

$q \supset r$

$\therefore\ p \supset r$

Consider the following argument:

1. John eats cereal and he runs a marathon and if John eats cereal, then he isn't hungry.
2. Either John is hungry or if John runs a marathon, then he is exhausted.
3. If John is exhausted, then he won't go to work.
4. Therefore, if John runs a marathon, then he won't go to work.

Let the following sentence constants stand in for the simple sentences above:

C = John eats cereal.

H = John is hungry.

M = John runs a marathon.

E = John is exhausted.

Given the above argument and sentence constants, we can construct the following derivation:

Derivation 10: Show $(M \supset \neg W)$

1.	$((C \wedge M) \wedge (C \supset \neg H))$	Premise
2.	$(H \vee (M \supset E))$	Premise
3.	$E \supset \neg W$	Premise
4.	$C \wedge M$	S, 1
5.	$C \supset \neg H$	S, 1
6.	C	S, 4
7.	$\neg H$	MP, 5 & 6
8.	$M \supset E$	DS, 2 & 7
9.	$M \supset \neg W$	HS, 8 & 3

12.2.8 Biconditional Elimination

Biconditional Elimination is an inference rule for the " $\equiv$ " operator. From any biconditional, one may derive a conditional whose antecedent is the left sentence connected by the biconditional operator and whose consequent is the right sentence connected by the biconditional operator, and vice versa. In schematic form, biconditional elimination is defined as:

$p \equiv q$

$\therefore\ p \supset q$

$\therefore\ q \supset p$

In other words, if a biconditional $p \equiv q$ is true, then $p \supset q$ is true and $q \supset p$ is true. Consider the following argument:

1. Quine is a philosopher if and only if John is.
2. Either John is a philosopher or Quine is a physicist.
3. If Quine is a physicist, then John loves Dewey, and if Quine is either a mathematician or a psychologist, then John doesn't love Dewey.

4. Quine is a mathematician and he enjoys etymology.
5. Therefore, Quine is a philosopher.

Let the following sentence constants stand in for the declarative sentences:

Q = Quine is a philosopher.

J = John is a philosopher.

P = Quine is a physicist.

D = John loves Dewey.

M = Quine is a mathematician.

G = Quine is a psychologist.

Given the above argument and sentence constants, we can construct the following derivation:

Derivation 11: Show Q

1.	$Q \equiv J$	Premise
2.	$J \vee P$	Premise
3.	$((P \supset D) \wedge ((M \vee G) \supset \neg D))$	Premise
4.	$M \wedge E$	Premise
5.	M	S, 4
6.	E	S, 4
7.	$P \supset D$	S, 3
8.	$((M \vee G) \supset \neg D)$	S, 3
9.	$M \vee G$	Add, 5
10.	$\neg D$	MP, 8 & 9
11.	$\neg P$	MT, 7 & 10
12.	J	DS, 2 & 11
13.	$J \supset Q$	BE, 1
14.	Q	MP, 12 & 13

12.2.9 Biconditional Introduction

Biconditional introduction is another inference rule for the "$\equiv$" operator. As mentioned in the section on logical operators, the biconditional is the conjunction of two conditionals. Thus, if it is such that $p \supset q$ is true and $q \supset p$ is true, then $p \equiv q$ is true. In schematic form, biconditional introduction is defined as:

$p \supset q$

$q \supset p$

$\therefore\ p \equiv q$

Recall the discussion on how to interpret conditional statements. The sentence $p \supset q$ can be read as "*p* only if *q*." We can also read the sentence $q \supset p$ as "*p* if *q*." Thus, we have "*p* if *q* and *p* only if *q*." Or, in other words, "*p* if and only if *q*."
Consider the following argument:

1. If Quine is a philosopher, then he enjoys etymology.
2. If Quine is a mathematician, then it is the case that if Quine enjoys etymology, then he is a philosopher.
3. Quine is either a mathematician or John loves Dewey, and John does not love Dewey.
4. Therefore, Quine is a philosopher if and only if he enjoys etymology.

Given the above argument and using the same sentence constants as before, we can construct the following derivation:

Derivation 12: Show $Q \equiv E$

1.	$Q \supset E$	Premise
2.	$(M \supset (E \supset Q))$	Premise
3.	$((M \vee D) \wedge \neg D)$	Premise
4.	$M \vee D$	S, 3
5.	$\neg D$	S, 3
6.	M	DS, 4 & 5
7.	$E \supset Q$	MP, 2 & 6
8.	$Q \equiv E$	BI, 1 & 7

The rules of inference discussed so far have all been inference rules for main logical operators. Simplification, conjunction, disjunctive syllogism, addition, modus ponens, modus tollens, hypothetical syllogism, biconditional elimination, and biconditional introduction *may only be used when the relevant operator is the main logical operator.*

12.3 Equivalence Rules

In the following sections, we introduce rules of replacement or equivalence rules. Rules of replacement *may be used on the relevant operator at any time in a derivation.* The rules of replacement that we will introduce are double negation (DN), commutation (COM), Association (ASSOC), contraposition (CONT), DeMorgan's theorem (DEM), and implication (IMP).

12.3.1 Double Negation (DN)

Double Negation is a replacement rule for the "¬" operator. Intuitively, it asserts that if it is not the case that it is not the case that some sentence is true, then that sentence is true. In other words, the denial of the denial of an assertion is equivalent to the assertion itself. In schematic form, double negation is defined as:

$\neg\neg p$	OR	p
$\therefore\ p$		$\therefore\ \neg\neg p$

In the practice of generating proofs, double negation allows for the introduction or elimination of two negation symbols prior to any well-formed formula in the derivation. In the following derivation, we find examples of repeated application of double negation:

Derivation 13: Show $\neg\neg(A \supset (\neg\neg\neg B \vee (C \wedge \neg\neg F)))$

1.	$(A \supset (\neg B \vee \neg\neg(\neg\neg C \wedge F)))$	Premise
2.	$(A \supset (\neg B \vee \neg\neg(\neg\neg C \wedge \neg\neg F)))$	DN, 1
3.	$(A \supset (\neg B \vee \neg\neg(C \wedge \neg\neg F)))$	DN, 2
4.	$(A \supset (\neg B \vee (C \wedge \neg\neg F)))$	DN, 3
5.	$(A \supset (\neg\neg\neg B \vee (C \wedge \neg\neg F)))$	DN, 4
6.	$\neg\neg(A \supset (\neg\neg\neg B \vee (C \wedge \neg\neg F)))$	DN, 5

12.3.2 Commutation (COM)

Commutation is a replacement rule based on the commutative properties of the "$\vee$" and "$\wedge$" functions. If $p \vee q$ is true, then so is $q \vee p$. Similarly, if $p \wedge q$ is true, then so is $q \wedge p$. In schematic form, commutation is defined as:

$p \vee q$	OR	$p \wedge q$
$\therefore q \vee p$		$\therefore q \wedge p$

For any conjunction or disjunction, the position of the conjuncts or disjuncts may be switched. Examples:

Derivation 14: Show $((D \vee C) \wedge (B \vee A))$

1.	$((A \vee B) \wedge (C \vee D))$	Premise
2.	$((C \vee D) \wedge (A \vee B))$	Com, 1
3.	$((D \vee C) \wedge (A \vee B))$	Com, 2
4.	$((D \vee C) \wedge (B \vee A))$	Com, 3

12.3.3 Association (ASSOC)

Association is another rule of replacement for the "$\vee$" and "$\wedge$" functions. It results from the associative property of these functions. For any sentence that is a uniform conjunction or a uniform disjunction, the position of the main logical function may be switched. In schematic form, Association is defined as:

$(p \wedge (q \wedge r))$	OR	$(p \vee (q \vee r))$
$\therefore ((p \wedge q) \wedge r)$		$\therefore ((p \vee q) \vee r)$

Examples of Association:

Derivation 15: Show $((((A \vee B) \vee C) \wedge D) \wedge E)$

1.	$((A \vee (B \vee C)) \wedge (D \wedge E))$	Premise
2.	$(((A \vee B) \vee C) \wedge (D \wedge E))$	ASSOC, 1
3.	$((((A \vee B) \vee C) \wedge D) \wedge E)$	ASSOC, 2

12.3.4 Contraposition (Cont)

Contraposition is a rule of replacement for the "⊃" operator. For any conditional statement, the antecedent and consequent may be switched provide that they are negated too. In schematic form, contraposition is defined as:

$p \supset q$	OR	$\neg p \supset \neg q$
$\therefore \neg q \supset \neg p$		$\therefore q \supset p$

Examples of contraposition:

Derivation 16: Show $(\neg(\neg C \supset \neg B) \supset \neg A)$

1.	$(A \supset (B \supset C))$	Premise
2.	$(A \supset (\neg C \supset \neg B))$	Cont, 1
3.	$(\neg(\neg C \supset \neg B) \supset \neg A)$	Cont, 2

12.3.5 Implication (Imp)

Implication is a rule of replacement for the "⊃" and "∨" operators. From any conditional, you may derive a disjunction whose left disjunct is the negation of the antecedent of that conditional and whose right disjunct is the consequent of that conditional, and vice versa. In schematic form, implication is defined as:

$p \supset q$	OR	$p \vee q$
$\therefore \neg p \vee q$		$\therefore \neg p \supset q$

Examples of implication:

Derivation 16: Show $(\neg(\neg A \vee B) \supset (\neg C \vee D))$

1.	$((A \supset B) \vee (C \supset D))$	Premise
2.	$((\neg A \vee B) \vee (C \supset D))$	Imp, 1
3.	$((\neg A \vee B) \vee (\neg C \vee D))$	Imp, 2
4.	$(\neg(\neg A \vee B) \supset (\neg C \vee D))$	Imp, 3

12.3.6 DeMorgan's Theorem (DEM)

DeMorgan's Theorem is a rule of replacement for the "∧" and "∨" operators. Given any conjunction or disjunction, you may derive a statement that replaces the disjunction operator with the conjunction operator or replaces a conjunction operator with a disjunction operator and inverts the values of all elements of the sentence and the sentence as a whole. In schematic form, DeMorgan's theorem is defined as:

$p \vee q$ $\therefore \neg(\neg p \wedge \neg q)$	OR	$p \wedge q$ $\therefore \neg(\neg p \vee \neg q)$
$\neg(p \vee q)$ $\therefore \neg p \wedge \neg q$	OR	$\neg(p \wedge q)$ $\therefore \neg p \vee \neg q$

Examples of DeMorgan's theorem:

Derivation 17: Show $\neg((\neg A \wedge \neg B) \vee \neg(\neg C \wedge \neg D))$

1.	$((A \vee B) \wedge \neg(C \vee D))$	Premise
2.	$(\neg(\neg A \wedge \neg B) \wedge \neg(C \vee D))$	DEM, 1
3.	$(\neg(\neg A \wedge \neg B) \wedge (\neg C \wedge \neg D))$	DEM, 2
4.	$\neg((\neg A \wedge \neg B) \vee \neg(\neg C \wedge \neg D))$	DEM, 3

It is interesting to note at this point the following. Given the rules of replacement above, we can show how to translate any sentence interpretable in the language of sentential logic into an equivalent sentence containing only two operators: negation and one other operator. This result is equivalent to what we saw in our investigation of normal forms in Chapter 11.

12.4 A Note on Reading Derivations in Sentential Logic

There are many different conventions for labeling proofs, but basically the purpose of these conventions is to allow readers to account for the legitimacy of each step. You can probably tell how the notational system for this proof works at this stage, but in virtually all derivational systems the following components will appear in one form or another.

The conclusion you're hoping to derive

What is asserted at each step?

What entitles you to make that assertion?

Show $-s$

The numbers serves as labels for each step so that you can identify them later when they serve as the basis for inferences

1.	$p \supset (s \supset m)$	Premise
2.	$p \wedge -m$	Premise
3.	p	S, 2
4.	$s \supset m$	MP 3 & 1
5.	$-m$	S, 2
6.	$-s$	MT 4 & 5

From line 2 by simplification

From lines 3 and 1 by Modus Ponens

12.5 Strategies, Proofs, and Formal Systems

At this point, we have shown how the rules of inference and the replacement rules are legitimate ways of connecting steps in reasoning. If an inference takes you from one step to another in a way that takes the same form as a rule of inference, or if it follows a replacement rule, then it is a logically legitimate piece of reasoning. As we introduced these rules in previous sections, we put them to work in derivations. These derivations were simple proofs designed to show how the conclusion followed from the premises of the proof.

When we approach the challenge of giving a derivation, we can adopt some general strategic principles. In this section, we will use the rules introduced so far in this chapter as the basis for constructing some proofs. These are the raw materials that will allow us to build a bridge from the premises to the conclusion. Using the rules introduced in the previous section, let's consider some examples of proof before talking a little more strictly about the system we are using. First, let's say someone started with the following three premises:

$p \supset (s \supset m)$
$(n \wedge -r) \wedge (m \supset t)$
p $\vee$ r

and they wanted to know whether these three premises entailed the conclusion:

$(s \supset t)$

With a little ingenuity, we could use the rules of inference from the previous section to derive a proof for the conclusion from the premises. As we attempt to construct what is called a direct proof for the validity of this argument, we can adopt the following strategy:

Step 1:

- Analyze the conclusion to determine where, or whether, the conclusion is stated as a part of the premises.
- If the conclusion is stated as part of the premise, then the task is to determine how it can stand on a line by itself after we apply the rules of inferences to the premises.
- Alternatively, if the whole conclusion doesn't seem to figure in the premises (this one doesn't). Then the task is to see whether the parts of the conclusion figure in the premises in a way that allows the derivation of the conclusion using replacement rules or rules of inference.

For instance in this case, we should look to see whether and where the *s* and *t* figure in the premises. They do, but not together.

$p \supset (\boldsymbol{s} \supset m)$

$(n \wedge -r) \wedge (m \supset \boldsymbol{t})$

$p \vee r$

Where does "$(s \supset t)$" show up in the premises?
The entire statement does not figure in the premises, but its parts do?

Step 2:

- Once we locate the conclusion or its parts in the premises, our analysis should move to the consideration of the main logical operators of each premise

$p \supset (\boldsymbol{s} \supset m)$

$(n \wedge -r) \wedge (m \supset \boldsymbol{t})$

$p \vee r$

What is the MLF of each premise and what rules of inference are available for formulas of this kind

The first premise has a $\supset$ as its MLF, so we should consider all the rules of inferences which apply to formulas of this kind.

The second premise has a.

The third premise has a $\vee$

- Once you see what the MLF is and have thought about the rules of inference that are available for sentences of this form, you should be thinking about ways of breaking the relevant parts of the conclusion out of the more complex sentences. Notice what's going on in each of the boxes below.

$p \supset (\boldsymbol{s} \supset m)$

$(n \wedge -r) \wedge (m \supset \boldsymbol{t})$

$\mathrm{p} \vee r$

In the first premise, you can think, for example, of
s
as trapped inside
$(s \supset m)$
which is trapped inside
$p \supset (s \supset m)$.

In the third premise, since the MLF is $\vee$, there are only a few options. Luckily, DS does the trick. Given $p \vee r$ if I also have $-r$, then I can conclude p. Getting p on a line by itself would be useful, since then I could use it to apply the rule of modus ponens such that given $p \supset (s \supset m)$ and p I can conclude $(s \supset m)$. That's one step closer to getting to the conclusion, but there are still some steps to go before we're there.

In the second premise, t is trapped inside
s
as trapped inside
$(m \supset t)$
which is trapped inside
$(n \wedge r) \wedge (m \supset t)$

You probably already noticed that since the MLF of the whole sentence is "and," it can be broken using the rule of simplification, leaving us with $(m \supset t)$ on a line by itself.

The boxes above are meant to represent the process of analysis that one might go through in the course of analyzing this problem. Remember, the goal is to provide a derivation of the conclusions; $(s \supset t)$ on a line by itself. Thus each of the boxes above gives part of the story,

Step 3:

- Build a gap free derivation of the conclusion based on the reasoning developed in Step 2. We will provide such a derivation below.

Show $(s \supset t)$

1.	$p \supset (s \supset m)$	Premise
2.	$(n \wedge -r)\ (m \supset t)$	Premise
3.	$p \vee r$	Premise
4.	$(n \wedge -r)$	S, 2
5.	$-r$	S, 4
6.	p	DS, 5 & 3

7.	$(s \supset m)$	MP, 6 & 1
8.	$(m \supset t)$	S, 2
9.	$(s \supset t)$	HS, 7 & 8

Working backwards from line 8 (the conclusion), we can ask how we derived that line from previous lines in the proof. We got line 9 "$(s \supset t)$" from lines 7 and 8 by hypothetical syllogism in the following manner:

$(s \supset m)$

$(m \supset t)$

therefore

$(s \supset t)$

HYPOTHETICAL SYLLOGISM (HS)

$p \supset q$

$q \supset r$

therefore

$p \supset r$

We will explain where line 8 came from in a moment, but for now, let's follow the ancestry of line 7.

$(s \supset m)$ also known as line 7 was derived from lines 6 and 1 by modus ponens

$p \supset (s \supset m)$

p

therefore

$(s \supset m)$

MODUS PONENS (MP)

$(p \supset q)$

p

therefore

q

Line 1 is a premise and so it needs no justification, but we do need to explain how we derived line 6. According to the proof, it was derived from 5 and 3 by disjunctive syllogism:

p v $\vee r$

$-r$

therefore

p

DISJUNCTIVE SYLLOGISM (DS)

$p \vee q$
$-p$
therefore
q

$p \vee q$
$-q$
therefore
p

Line 3 is a premise too, so it also needs no justification. Line 5 was derived from line 4 from simplification and line 4 was derived from line 2 by simplification. Line 2 is a premise, so it needs no justification.

The conclusion: line 9, was derived by hypothetical syllogism from lines 7 and 8. So far, we have shown the entire ancestry of line 7 from the all the way back to the premises. We have yet to do the same for line 8. But this is easy since line 8 was derived by simplification from line 2, and line 2 is a premise.

It is sometimes difficult to find a way to derive the conclusion from the premises using the rules of inference, and we can resort to other methods, as we will see in the following section. However, for now, let's practice with some examples.

12.5.1 Some Examples of Proofs and the Analyses that Underlay Them

In the following pages, we examine some examples of derivations using the replacement rules and rules of inference introduced in this chapter so far. Remember that the goal is to demonstrate the validity of each of the following arguments. In the problem, the numbered lines provided are the premises of the argument. The conclusion is marked by the "$\therefore$" symbol.

Problem 1:

1. $R \wedge S$
2. T

$\therefore ((T \vee L) \wedge (R \wedge S))$

Solution 1:

Show $((T \vee L) \wedge (R \wedge S))$

1.	$R \wedge S$	Premise
2.	T	Premise
3.	$T \vee L$	Add, 2
4.	$((T \vee L) \wedge (R \wedge S))$	Conj, 3 & 1

In this case, we do not find the whole conclusion $((T \vee L) \wedge (R \wedge S))$ in the premises, but we do find most of its parts. $R \wedge S$ is a premise on a line by itself and so it T. We know that by the rule of addition, we can add anything to a wff on a line by itself using disjunction, so if we have T, then we can conclude $\vee L$. Given that we can have $T \vee L$ and $R \wedge S$ on lines by themselves, we can assert their conjunction $((T \vee L) \wedge (R \wedge S))$ by the rule of conjunction.

In our next example, we will consider two different ways of running the derivation. Both are equally good.

Problem 2:

1. C
2. $C \supset A$
3. $(A \supset (B \wedge D))$
$\therefore B$

FIRST SOLUTION

Show B

1.	$C \supset A$	Premise
2.	$(A \supset (B \wedge D))$	Premise
3.	C	Premise
4.	A	MP, 1 & 3
5.	$B \wedge D$	MP, 2 & 4
6.	B	S, 5

Our first step is to locate the conclusion, in this case B. We see that B is trapped inside a conjunction $(B \wedge D)$ that is itself trapped inside a conditional. $(B \wedge D)$ is the consequent of the conditional $(A \supset (B \wedge D))$. In order to get $(B \wedge D)$ on a line by itself, we need to locate the antecedent of the conditional; A . A is in the second premise and is also trapped inside a conditional $C \supset A$. In order to get A on a line by itself, we need to locate the antecedent of the conditional; C. Happily C is sitting on a line by itself, as our first premise. Thus, given C, we can get A by Modus Ponens with the second premise, then we can use A that we derive in that way to derive B by Modus Ponens with the third premise.

But there are many ways to build our derivation from the premises to the conclusion. Let's examine an alternative to the first solution that uses hypothetical syllogism instead of a pair of applications of modus ponens:

SECOND SOLUTION

Show B Exercise 7:

Show B

1.	$C \supset A$	Premise
2.	$(A \supset (B \wedge D))$	Premise
3.	C	Premise
4.	$(C \supset (B \wedge D))$	HS, 1 & 2
5.	$B \wedge D$	MP, 3 & 4
6.	B	S, 5

At this point, let's examine how rules of inference like hypothetical syllogism require us to find and exploit patterns in premises in order to generate elegant proofs.

Problem 3:

1. $((B \wedge M) \supset R)$
2. $(L \supset (B \wedge M))$
$\therefore L \supset R$

Solution 2:

Show $L \supset R$

1.	$((B \wedge M) \supset R)$	Premise
2.	$(L \supset (B \wedge M))$	Premise
3.	$L \supset R$	HS, 1 & 2

This is another case where we do not find the whole conclusion $L \supset R$ in the premises. We do find both L and R in the premises and they are both parts of conditionals. Our second step is to examine the premises to determine their main logical function before thinking about which rules of inference we can deploy with respect to those functions. We should also notice structural features of the premises. For example, the antecedent of the first conditional is identical to the consequent of the second conditional. Cases like this should make us think of the rule of hypothetical syllogism. In the following proof, hypothetical syllogism plays a crucial role again:

Problem 5:

1. $(A \supset (A \wedge B))$
2. $C \supset A$

$\therefore ((C \supset (A \wedge B)) \wedge (C \supset A))$

Solution 5:

Show $((C \supset (A \wedge B)) \wedge (C \supset A))$

1.	$(A \supset (A \wedge B))$	Premise
2.	$C \supset A$	Premise
3.	$(C \supset (A \wedge B))$	HS, 1 & 2
4.	$((C \supset (A \wedge B)) \wedge (C \supset A))$	Conj, 3 & 2

<u>Some Exercises and Sample Solutions:</u>

Exercise 1:

1. $R \vee S$
2. $((A \supset L) \wedge ((R \vee S) \supset T))$
$\therefore T \vee L$

Exercise 2:

1. $A \wedge B$
2. $B \supset C$
$\therefore C$

Exercise 3:

1. $A \supset B$
2. $C \wedge A$
$\therefore B \vee D$

Exercise 4:

1. $(A \supset (\neg B \wedge C))$
2. $C \supset D$
3. $E \vee B$
4. A
$\therefore D \wedge E$

Exercise 5:

1. $((F \supset G) \vee H)$
2. $\neg G$
3. $\neg H$
$\therefore \neg F$

Exercise 6:

1. L
2. $T \vee \neg R$
3. $((L \vee R) \supset \neg T)$
$\therefore \neg R \vee B$

Exercise 7:

1. $((R \wedge A) \vee E)$
2. $((R \wedge A) \supset D)$
3. $\neg D$
$\therefore E \wedge \neg D$

Exercise 8:

1. $((A \wedge D) \supset \neg C)$
2. $((R \vee S) \supset (A \wedge D))$
3. $(\neg C \supset \neg(A \wedge D))$
$\therefore ((R \vee S) \supset \neg(A \wedge D))$

Exercise 9:

1. $((A \vee \neg C) \supset B)$
2. A
3. $((A \vee \neg D) \supset (R \wedge S))$
$\therefore ((R \wedge S) \wedge B)$

Exercise 10:

1. $\neg A$
2. $((C \vee A) \supset L)$
3. $A \vee D$
4. $((D \vee U) \supset C)$
$\therefore L$

Exercise 11:

1. $(R \supset (\neg P \vee \neg M))$
2. $(\neg R \supset (\neg M \wedge \neg N))$
3. $\neg(\neg P \vee \neg M)$
4. $Z \vee R$
$\therefore ((\neg M \wedge \neg N) \wedge Z)$

Exercise 12:

1. A
2. $((B \vee C \supset D)$
3. $((A \vee E) \supset (B \wedge C))$
$\therefore D$

Exercise 13:

1. $A \vee B$
2. $C \supset A$
3. $((B \wedge \neg C) \supset (D \wedge \neg C))$
4. $\neg A$
$\therefore D$

Exercise 14:

1. $((\neg A \wedge \neg B) \supset (C \supset B))$
2. $B \supset A$
3. $\neg A$
$\therefore \neg C$

SOME SOLUTIONS:

Exercise 5:
Show $\neg F$

1. $((F \supset G) \vee H)$	Premise
2. $\neg G$	Premise
3. $\neg H$	Premise
4. $F \supset G$	DS, 1 & 3
5. $\neg F$	MT, 2 & 4

Exercise 7:

Show $E \wedge \neg D$

1.	$((R \wedge A) \vee E)$	Premise
2.	$((R \wedge A) \supset D)$	Premise
3.	$\neg D$	Premise
4.	$\neg(R \wedge A)$	MT, 2 & 3
5.	E	DS, 1 & 4
6.	$E \wedge \neg D$	Conj, 5 & 3

Exercise 9:

Show $((R \wedge S) \wedge B)$

1.	$((A \vee \neg C) \supset B)$	Premise
2.	A	Premise
3.	$((A \vee \neg D) \supset (R \wedge S))$	Premise
4.	$A \vee \neg D$	Add, 2
5.	$R \wedge S$	MP, 3 & 4
6.	$A \vee \neg C$	Add, 2
7.	B	MP, 1 & 6
8.	$((R \wedge S) \wedge B)$	Conj, 5 & 7

Exercise 11:

Show $((\neg M \wedge \neg N) \wedge Z)$

1.	$(R \supset (\neg P \vee \neg M))$	Premise
2.	$(\neg R \supset (\neg M \wedge \neg N))$	Premise
3.	$\neg(\neg P \vee \neg M)$	Premise
4.	$Z \vee R$	Premise
5.	$\neg R$	MT, 1 & 3
6.	$\neg M \wedge \neg N$	MP, 2 & 5
7.	Z	DS, 4 & 5
8.	$((\neg M \wedge \neg N) \wedge Z)$	Conj, 6 & 7

Exercise 13:

Show D

1.	$A \vee B$	Premise
2.	$C \supset A$	Premise
3.	$((B \wedge \neg C) \supset (D \wedge \neg C))$	Premise
4.	$\neg A$	Premise
5.	$\neg C$	MT, 2 & 4
6.	B	DS, 1 & 4
7.	$B \wedge \neg C$	Conj,- 6 & 5
8.	$D \wedge \neg C$	MP, 3 & 7
9.	D	S, 8

13 Conditional Proof and Proof by Contradiction in Sentential Logic

13.1 The Limits of Direct Proof

The previous chapter explained how to demonstrate the validity of some arguments by showing how the conclusion can be derived in a step-by-step manner from the premises via the application of rules. The derivations we studied are examples of what is known as *direct proof*. However, it is not always possible to generate direct proofs to demonstrate the validity of an argument.

Chapter 12 described how direct proofs are restricted to using premises, rules of inference, and replacement rules. Given this restriction it is simply not possible to demonstrate the validity of some arguments for which we could easily generate truth table proofs of the kind we learned in Chapter 11. Thus, we know that there are some valid arguments for which no direct proof can be given.

Take the claim that a logically true sentence, or a tautology, (e.g., $q \vee \neg q$) follows from any claim whatsoever. We know that this is correct thanks to the truth table method, but let's say we encounter an argument such as the following:

R
$\therefore Q \vee \neg Q$

Here, we an argument claiming that from the premise R we can conclude $Q \vee \neg Q$. We know that any logically true sentence follows from any sentence, but how would we begin to prove that this argument is valid using direct proof? Given our strategic advice in Chapter 12 there seem to be no options available. Recall that Step 1 of our strategic advice for constructing direct proofs involved analyzing the conclusion to determine where, or whether, the conclusion appears as a part of the premises. In direct proof, if the conclusion is stated as part of the premise, then the task is to determine how it can stand on a line by itself after we apply the rules of inferences to the premises. Since the first step in our strategy fails in this case we will need to explore other options. We know that that truth table method offers a means to prove that this is a valid argument; we can easily show that this argument, when presented in the form of a conditional, is a logical truth or a tautology.

R	⊃	(*Q*	∨	¬	*Q*)
T	T	T	T	F	T
T	T	F	T	T	F
F	T	T	T	F	T
F	T	F	T	T	F

The values falling under the main logical function ⊃ are all true. Therefore, the sentence as a whole is true under all circumstances. However, the reason we turned to direct proofs in Chapter 12 was because of the cumbersome nature of the truth table method for more complicated cases. In a case like this, for instance, we can probably see that it is obvious that the sentence is a tautology and it seems unnecessarily time consuming to have to lay out an entire truth table to demonstrate it. Happily, we also have two additional methods of proof; conditional and proof by contradiction, to which we can turn. These are common methods of proof in mathematical reasoning. In practice, they both rely on introducing assumptions that are not given in the premises in order to generate proofs of the validity of arguments. This chapter explains how these proofs work and justifies the introduction of assumed premises.

To begin with, conditional proof and proof by contradiction both rely on characteristics of our definition of the material conditional. Recall that in tabular form, we define "⊃" as:

p	$\supset$	q
T	T	T
T	F	F
F	T	T
F	T	F

The conditional is true in the first, third, and fourth rows. It is false only in the second row. The second row represents the situation in which p is true and q is false. Thus, we know that whenever p is false, the conditional as a whole is true. p is false in the third and fourth rows only, and according to our definition of the $\supset$ function, the sentence as a whole is true. This definition of the conditional can be exploited to provide us with our two new methods; conditional proof and proof by contradiction.

13.2 Conditional Proof

Let's first introduce conditional proof. **In a conditional proof, the goal is show that we can conclude some conditional follows validly from the premises or that the conditional itself is a tautology**.

At this point, we are ready to quickly sketch the reasoning involved in the method of conditional proof:

- We are trying to prove that some conditional follows validly from the premises of an argument or that the conditional is itself a tautology or a logical truth.
- Let's think about the parts of the conditional we are trying to prove. First, let's think about the antecedent.
- Maybe *the antecedent* (p) is true, maybe it isn't. We don't know. But given the principle of bivalence we know that it is either true or false. Let's examine both options.
- We know that if *the antecedent* (p) is false then *the conditional* ($p \supset q$) as a whole is true. To understand why, we just need to look back at the definition of the conditional (look at rows 3 and 4 on the truth table).
- Given that the only options remaining are the first and second row on the truth table, then in order to determine whether the conditional as a whole is true we only need to consider the situations in which the antecedent is true.

- We ask ourselves what happens when *the antecedent* is true? How are we going to know what happens when the antecedent is true?
- Well, let's just assume that *the antecedent* is true and see what happens.
- If we assume for the sake of argument that the antecedent p is true and it turns out that *the conditional* as a whole *must be true* or *follows logically from* the truth of the antecedent then things get interesting.
- If we can prove that *the conditional* as a whole *must be true*, when we assume that the antecedent is true
  and
- given the fact that if *the antecedent* is false then we know that *the conditional* ($p \supset q$) as a whole *must be true*
  and
- since the antecedent can only be true or false
  and
- in both cases the conditional as a whole is true
  therefore
- the conditional as a whole *must be* true.

Conditional proof and proof by contradiction involve making an assumption. Specifically, we introduce a new premise, a so-called *assumed premise*, to the argument. In a conditional proof, **we assume the truth of the antecedent of the conditional that we want to demonstrate** given the premises of the argument. As we shall see below, a different kind of assumed premise will be introduced in the case of proof by contradiction. If the consequent follows from the premises of an argument and the assumption of the truth of the antecedent of the conditional one is trying to prove, then we can validly conclude the conditional.

13.2.1 Introducing the technique for conditional proofs

Imagine being asked to prove that

$$Q \supset (Q \vee \neg Q)$$

is logically true. How would we proceed?

The truth table method is likely the first place we would turn. We can demonstrate that under all circumstances the statement as a whole is true. However, we can now see how a conditional proof could also work in this context. If we assumed the antecedent of the conditional; Q we know that by the rule of addition we can connect $\neg Q$ to it with a disjunction. Thus, assuming Q we can conclude $Q \vee \neg Q$. In our system, we will need to carefully account for and document the introduction of assumed premises. Proof by contradiction and conditional proofs introduce a *subproof* into a derivation. Subproofs begin by making an assumption that is relevant to particular kind of subproof one wants to conduct. It is important to keep track of your assumptions. In order to do so, we indent the line on which an assumption is made and add subproof numberings (1.1, 1.2, etc.). When constructing a derivation, all subproofs must close. We mark the closing of a conditional proof by stating the conditional shown on a new line and using the "CP" justification. Once an assumption has been closed, the subproof lines of the derivation can no longer be used in the rest of the proof.

Let's consider an example that calls for the introduction of an assumed premise for the sake of a conditional proof.

$(A \supset (B \vee \neg C))$

$(C \wedge (B \supset D))$

$\therefore (A \supset D)$

Show $(A \supset D)$

1.	$(A \supset (B \vee \neg C))$	Premise
2.	$(C \supset (B \supset D))$	Premise
2.1	A	Assumption CP
2.2	$B \vee \neg C$	MP, 1 & 2.1
2.3	C	S, 2
2.4	B	DS, 2.2 & 2.3
2.5	$B \supset D$	S, 2
2.6	D	MP, 2.4 & 2.5
3.	$(A \supset D)$	CP, 2.1–2.6

On 2.1, we introduce *A* as an assumed premise for the sake of a conditional proof. We indent and number the lines of the subproof accordingly.

On 2.6, we have derived the consequent of the antecendent; *D* using the assumed premise *A*. This means that given *A*, *D* follows. Line 3 is justified by the work done in the subproof from lines 2.1–2.6.

Notice that once we introduce the assumed premise, we indent the proof and start numbering the subproof accordingly. Once the subproof is complete, we return to the main proof.

This proof required the introduction of the assumed premise A for the sake of a conditional proof that $A \supset D$. At 2.1, with the introduction of the assumed premise, our accounting method calls for us to indent the proof and start numbering the subproof accordingly. Once the subproof is complete, we return to the main proof. Occasionally, it will be necessary to conduct multiple subproofs. In the next example, we will see a case where more than one assumed premise is introduced and a subproof occurs within a subproof. Notice that in the following example, two conditionals proofs one conditional proof will not be used to demonstrate the conclusion of the argument directly. In fact, conditional proofs may be used to prove any conditional anywhere in a derivation. Take a look at a proof for the following argument:

$(A \supset (B \supset C))$

$(B \supset (C \supset D))$

$\therefore (A \supset (B \supset D))$

In this case, we will need two subproofs and we will need to introduce two assumed premises:

Show $(A \supset (B \supset D))$

1. $(A \supset (B \supset C))$	Premise
2. $(B \supset (C \supset D))$	Premise
2.1 A	Assumption CP
2.2 $B \supset C$	MP, 1 & 2.1
2.2.1 B	Assumption CP
2.2.2 $C \supset D$	MP, 2 & 2.2.1
2.2.3 C	MP, 2.2 & 2.2.1
2.2.4 D	MP, 2.2.2 & 2.2.3
2.3 $B \supset D$	CP, 2.2.1–2.2.4
3. $(A \supset (B \supset D))$	CP, 2.1–2.3

Let's look at one more example of how we can use conditional proofs as a component of a larger proof.

$(A \supset D) \supset \neg B))$

$(A \supset (C \wedge D))$

$\therefore \neg B \vee F$

In this case, we are not trying to derive a conditional as the conclusion, but we will need to use a conditional $(A \supset D)$ in order to derive $\neg B$. Once we have $\neg$, we can derive $\neg B \vee F$ by the rule of addition.

Show $\neg B \vee F$

1. $(A \supset D) \supset \neg B$	Premise
2. $(A \supset (C \wedge D))$	Premise
2.1 A	Assumption CP
2.2 $C \wedge D$	MP, 2 & 2.1
2.3 D	S, 2.2
3. $(A \supset D)$	CP, 2.1–2.3
4. $\neg B$	MP, 1 & 3
5. $\neg B \vee F$	ADD, 4

13.3 Proof by Contradiction or *Reductio Ad Absurdum* (RAA)

Proof by contradiction also known as *reductio* proofs (short for *reductio ad absurdum*) is one of the most powerful methods available to us in formal reasoning. Some mathematicians and philosophers, most notably L. E. J Brouwer worried that the *reductio* method of proof is simply *too* powerful and that it rests on a faulty assumption[1]. Brouwer rejected the assumption known as the principle of the excluded middle. This principle, like the principle of bivalence, states that a declarative sentence is either true or false, there is no third option.

[1] Brouwer, L. E. (1907). On the foundation of mathematics. *Collected Works*, *1*, 11–101.

If the principle of excluded middle holds, then a mathematician who wants to prove that a given sentence is true, say some claim about the nature of geometrical objects, sets, groups, or numbers, can generate a proof by introducing an assumed premise that denies the claim. From there, the mathematician must show that this assumed premise leads to a contradiction. Given the principle of the excluded middle, if the negation of a claim leads to a contradiction then we have demonstrated that the positive assertion of the claim must hold.

The power of this method of proof is quite striking. For example, let's say you want to show that the square root of two cannot be expressed as the ratio of two numbers (that it is irrational). One way to prove this is by showing that the negation of what we are trying to prove leads to a contradiction. Assuming that the square root of two is rational generates a contradiction, therefore the square root of two must be irrational. For fun, let's see what such a proof would look like.

Assume that $\sqrt{2}$ is rational means that it can be written as a ratio of two numbers p and q with no common denominator as follows:

$$\sqrt{2} = \frac{p}{q}$$

Now, squaring both sides gives

$$2 = \frac{p^2}{q^2}$$

This means that

$$p^2 = 2q^2$$

Notice that p^2 is even. This means that p is even. p^2 must be divisible by 4 and if this were the case then q^2 and hence q must be even. We must conclude that p and q are even in violation of our assumption that they have no common denominator. This contradicts our assumption. We cannot accept a contradiction, therefore the assumption that $\sqrt{2}$ is rational must be rejected. If the claim that $\sqrt{2}$ is rational is a contradiction, then the denial of that claim is a tautology. The negation of this assumption is that $\sqrt{2}$ is irrational. Therefore, $\sqrt{2}$ is irrational.[2]

[2] I owe this version of the proof to Professor Peter Alfeld at the University of Utah. See his *Understanding Mathematics: A Study Guide* available at http://www.math.utah.edu/~pa/math.html (last accessed July 7, 2017).

All across mathematics, proof by contradiction is deployed in a variety of creative and important ways.

13.3.1 Introducing the technique for conditional proof by contradiction in sentential logic

We will follow the same method for annotating proofs by contradiction as described above for conditional proofs. When we assume the denial of the conclusion for the sake of a proof by contradiction we will introduce it as an assumed premise. The subproof will be indented in the same way as previously introduced. In the subproof for a proof by contradiction the goal will be to derive a contradiction. Once the contradiction is derived, the subproof ends and the conclusion can be asserted as on a line following.

The general form of a proof by contradiction is as follows:

Show p

1.	**. . .**	Premise
1.1.	$\neg p$	Assumption RAA
	.	
	.	
	.	
1.2.	q	. . .
	.	
	.	
	.	
1.3.	$\neg q$	. . .
1.4.	$q \wedge \neg q$	Conj, 1.2 & 1.3
2.	p	RAA, 1.1–1.4

In the above schema, we began by assuming the negation of what we were trying to prove. This assumption allowed us to derive a contradiction (Line 1.4). All subproofs that function as proofs by contradiction must close with a contradiction. If an assumption leads to a contradiction, then the inverse of the assumption follows in the derivation as in line 2.

Consider the following derivation using indirect proof:

Show A	
1. $A \vee B$	Premise
2. $A \vee \neg B$	Premise
2.1. $\neg A$	Assumption RAA
2.2. B	DS, 1 & 2.1
2.3. $\neg B$	DS, 2 & 2.1
2.4. $B \wedge \neg B$	Conj, 2.2 & 2.3
3. A	RAA, 2.1–2.4

All assumptions must close. A derivation is not complete if there are any open assumptions. We mark the closure of an indirect proof with a line containing what was shown by the indirect proof and the justification "RAA" (Line 3 above). Once an assumption has been closed, all lines that were part of the assumption are no longer usable in the derivation.

Here are some more examples of proofs by contradiction:

$(A \supset B) \wedge C$
$(A \supset \neg B) \wedge D)$
$\therefore \neg A$

Show $\neg A$	
1. $(A \supset B) \wedge C$	Premise
2. $(A \supset \neg B) \wedge D$	Premise
2.1. $\neg\neg A$	Assumption RAA
2.2. A	DN, 2.1
2.3. $(A \supset B)$	S, 1
2.4. $(A \supset \neg B)$	S, 2
2.5. B	MP, 2.2 & 2.3
2.6. $\neg B$	MP, 2.2 & 2.4
2.7. $B \wedge \neg B$	CONJ, 2.5 & 2.6
3. $\neg A$	RAA, 2.1–2.7

$(A \vee B) \vee (C \wedge B)$
$(A \supset B) \wedge \neg C$
$(F \supset ((A \supset B) \wedge \neg C))$
$\therefore (A \vee B)$

Show $(A \vee B)$

1. $(A \vee B) \vee (C \wedge B)$	Premise
2. $(A \supset B) \wedge \neg C$	Premise
3. $(F \supset ((A \supset B) \wedge \neg C))$	Premise
3.1. $\neg (A \vee B)$	Assumption RAA
3.2. $C \wedge B$	DS, 3.1 & 1
3.3. C	S, 3.2
3.4. $\neg C$	S, 2
3.5. $C \wedge \neg C$	CONJ, 3.3 & 3.4
4. $(A \vee B)$	RAA, 3.1–3.5

It is often necessary to generate more than one subproof in the course of an argument. For example, in the following proof it is necessary to derive the antecedent of the conditional in the first premise and the denial of one of the disjuncts in the first premise. Two assumptions are introduced in the following proof in order to derive the needed pieces of the proof.

$(\neg A \supset (E \vee F)) \vee M$
$\neg C \wedge \neg M$
$(A \supset B) \wedge (E \supset B)$
$(A \supset \neg B) \wedge (E \supset \neg B)$
$\therefore (F \vee N)$

Show $(F \vee N)$

1. $(\neg A \supset (E \vee F)) \vee M$	Premise
2. $\neg C \wedge \neg M$	Premise
3. $(A \supset B) \wedge (E \supset B)$	Premise
4. $(A \supset \neg B) \wedge (E \supset \neg B)$	Premise
5. $\neg M$	S, 2
6. $(A \supset B)$	S, 3
7. $(A \supset \neg B)$	S, 4
7.1. A	Assumption RAA
7.2. B	MP, 7.1 & 6
7.3. $\neg B$	MP, 7.1 & 7
8. $\neg A$	RAA, 7.1–7.3

9. $(\neg A \supset (E \vee F))$	DS, 1 & 5
10. $(E \vee F)$	MP, 8 & 9
10.1. E	Assumption RAA
10.2. $(E \supset B)$	S, 3
10.3. $(E \supset \neg B)$	S,4
10.4. B	MP, 10.1 & 10.2
10.5. $\neg B$	MP, 10.1 & 10.2
11. $\neg E$	RAA, 10.1–10.5
12. F	DS, 10 & 11
13. $(F \vee N)$	ADD 12

14 Proofs with Trees

14.1 Introducing the Tree Rules

Truth tables and deductive proofs provide two ways to determine the validity of an argument. Truth tables are relatively cumbersome in the context of more complicated arguments. We turned to systems of proof that allow us to derive the conclusion from the premises by means of well-justified rules of inference. Deductive proofs of the kind we studied in Chapters 12 and 13 sometimes have the disadvantage of requiring considerable ingenuity to generate. Moreover, yet, we know from the truth table method that the question of validity is decidable for sentential logic. In this section, we will introduce another method that we can use to decide the validity of arguments without any ingenuity. This is the so-called tree method. The method is simple and mechanical and does not require us to come up with clever proofs. It is also efficient insofar as it eliminates the redundancy and repetition of complicated truth tables. At its heart, the tree method is basically equivalent to the truth table method and the trees are sometimes called abbreviated truth tables.

One way of thinking about the tree method is that it provides a different kind of graphical representation of the truth table method. The main features of this graphical representation are paths and branches.

The tree method works by taking well-formed formulas (wffs) and breaking them into their constituent parts according to rules. We are familiar with how these rules are justified. They way that the rules are applied looks slightly different in this context. Let's begin with the simplest case. Given the conjunction:

$$(A \wedge B)$$

we know that we can legitimately conclude both A by itself and B by themselves. This was the rule of simplification that we encountered previously in our study of derivations. In our trees, we represent the application of the rule in the following way:

$$\begin{array}{c} (A \wedge B) \\ | \\ A \\ B \end{array}$$

At the top of the tree is the well-formed formula, in this case the conjunction of A and B. Falling underneath the conjunction is a vertical line representing a path. In the path falling under the conjunction are the two conjuncts listed separately. The path represents our reasoning; we know that if the conjunction is true, both conjuncts must also be true. By contrast, consider the disjunction

$$(A \vee B)$$

In this case, we know from our truth table representation of the disjunction that either A is true, B is true, or both A and B are true. We have three kinds of scenario in which the sentence as a whole is true and one where it is false.

p	$\vee$	q
T	T	T
T	T	F
F	T	T
F	F	F

We represent this state of affairs by picturing our situation in terms of a fork in the tree. The fork should be interpreted as offering three possibilities, the one in which A is true, the one in which B is true, and given that there are two open paths, the third possibility is that A and B are both true.

$$(A \vee B)$$

$$A \qquad\qquad B$$

The tree contains two open paths falling under the disjunction.

The first path should be read as containing $(A \vee B)$ and A. The second should be read as containing $(A \vee B)$ and B. The fact that both open paths are present means that the tree also includes the possibility where $(A \vee B)$ and A ***and*** $(A \vee B)$ and B. Notice that both paths also contain A ∨ B. When we say that two paths are open, we mean that the open paths represent the ways that things could be.

Now let's say we have a slightly more complicated formula as follows:

$$(A \vee B) \wedge \neg B$$

We can apply the two tree rules that we have seen so far. We have a rule for $\wedge$ and a rule for $\vee$. We can only apply the tree rule to the MLF of the sentence which in this case is the $\wedge$.

$$(A \vee B) \wedge \neg B$$

$$(A \vee B)$$

$$\neg B$$

By applying the tree rule for $\wedge$ to the MLF of $(A \vee B) \wedge \neg B$, we create a path in which both conjuncts $(A \vee B)$ and $\neg B$ fall. Notice that we can apply the tree rule for the disjunction to $(A \vee B)$. If we do this, we will generate the following tree:

$$(A \vee B) \wedge \neg B$$

$$(A \vee B)$$

$$\neg B$$

$$A \qquad B$$

Falling under the $\neg B$, we list the result of applying the tree rule for disjunction to the MLF of $(A \vee B)$. This gives us a tree consisting of two paths:

The path first reads $\quad (A \vee B) \wedge \neg B,\ (A \vee B),\ \neg B,\ A$

The path second reads $\quad (A \vee B) \wedge \neg B, (A \vee B),\ \neg B,\ B$

Notice that the second path contains a contradictory pair of statements; $\neg B$, B. Because of this, we know that the second path is not a possible way that things could be and we are therefore entitled to ignore it. We call a path containing a contradiction a **closed path.** In this case, we have a tree with only one open path, namely the first.

Now consider the following well-formed formula:

$$\neg (A \vee B)$$

Here the MLF is neither a disjunction nor a conjunction, but rather the negation. In order to explain the tree rule for the negation of a disjunction, we will need to return again to our truth tables. When we consider the truth table for this statement below,

¬	(A	∨	B)
F	T	T	T
F	T	T	F
F	F	T	T
T	F	F	F

The truth table can help us to see that —(A ∨ B) is only true in case A is false and B is false (the fourth case on the truth table).

we see that the total value for the well formed is such that it is only true when A is false and B is false. This is the fourth row on the truth table. Given this, we know that

$$\neg (A \vee B)$$

is logically equivalent to

$$\neg A \wedge \neg B$$

Thus, the negation of a disjunction is logically equivalent to negating each disjunct and joining both together with the conjunction. The tree rule for the negation of the disjunction is simply a single path containing the negation of both disjuncts:

$$\neg (A \vee B)$$
$$|$$
$$\neg A$$
$$\neg B$$

By the same reasoning, we can see that the negation of the conjunction will involve two open paths each containing the negation of one of the conjuncts. This is because the negation of a conjunction is equivalent to a disjunction of the negation of both conjuncts. That is a very cumbersome way of saying that we know by DeMorgan's law that $\neg (A \wedge B)$ is equivalent to $\neg A \vee \neg B$. Notice that in this case, we are taking equivalent wffs that have a disjunct as the MLF rather than the negation. Once we can see that $\neg (A \wedge B)$ can be restated as an equivalent sentence with a disjunction as the MLF, then we can apply the tree rule for disjunction as follows:

$$\neg (A \vee B)$$
$$/ \quad \backslash$$
$$\neg A \qquad \neg B$$

Given the close connection between the truth tables on the tree rules, we can determine the tree rule for the horseshoe. Remember that a well-formed formula containing the horseshoe as its main logical operator is only false when the antecedent is true and the consequences are false. Thus, we know that asserting

$$\mathrm{A} \supset \mathrm{B}$$

is equivalent to asserting the denial of the second row on the truth table

(A	⊃	B)
T	T	T
T	F	F
F	T	T
F	T	F

The second row on the truth table is the only case where A ⊃ B is false.

The second row is the case where A ∧ – B
Thus,
A ⊃ B
is equivalent to
–(A ∧ – B)

$$\neg(A \wedge \neg B),$$

as we saw above, is equivalent to

$$\neg A \vee \neg\neg B$$

by DeMorgan's law. We also know that

$$\neg\neg B$$

is logically equivalent to

$$B$$

so we know that

$$A \supset B \text{ is equivalent to } \neg A \vee B.$$

Now, given that we know that paths split in the case of disjunction, we can finally state our tree rule for material conditionals:

$$(A \supset B)$$

$$\neg A \qquad \neg B$$

We can devise a general recipe for representing any operator in terms of paths on the tree. We recall that we were able to convert any statement of sentential logic into a normal form consisting of negations, conjunctions, and disjunctions. This means that whenever any operator appears as the MLF of a formula, that formula can be restated in tree form by applying the tree rules that we have already learned for disjunction or conjunction to an equivalent statement of the formula in normal form.

Put more simply, the recipe for figuring out the tree rule for some operator runs as follows: First, examine the truth table for the operator. Find some equivalent form to the operator that uses only $\wedge$, $\vee$, and $\neg$. Once we have this normal form version of the formula, it is a straightforward matter to apply the tree rules that we have learned so far. For example, in the case of the negation of the material conditional:

¬	(A	⊃	B)
F	T	T	T
T	T	F	F
F	F	T	T
F	F	T	F

The second row on the truth table is the only case where A ⊃ B is false. This means that in the case of the negation of A ⊃ B this will be the only case that is true.

The second row is the case where A ∧ –B
Thus,
–(A ⊃ B)
is equivalent to (A ∧ –B)
–(A ∧ –B)

Given that ¬ (A ⊃ B) is logically equivalent to (A ∧ ¬B) the tree rule for a formula of this kind is:

$$\neg\,(A \supset B)$$

$$|$$

$$\begin{array}{c} \mathrm{A} \\ \neg \mathrm{B} \end{array}$$

Following the same general recipe, we can determine the tree rules for other operators.

Recall from Chapter 11, the exhaustive list of the 16 binary operators:

0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
F	F	F	F	F	F	F	F	T	T	T	T	T	T	T	T
F	F	F	F	T	T	T	T	F	F	F	F	T	T	T	T
F	F	T	T	F	F	T	T	F	F	T	T	F	F	T	T
F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T

When the biconditional is the MLF, for example, we recall that it is equivalent to number 1001 or

1001 ≡
T
F
F
T

This means that the formula is true in the first or fourth cases: The case where A is true and B is true or the cases where $\neg A$ is true and $\neg B$ is true. This tells us that in normal form, we can read the following equivalent statement from the truth table:

$$((A \wedge B) \vee (\neg A \wedge \neg B))$$

It is a straightforward matter to apply the two tree rules that we know first to the MLF (the disjunction) and then to the MLF of the two disjuncts as follows:

$$(A \equiv B)$$

$$\begin{array}{cc} A & \neg A \\ B & \neg B \end{array}$$

Let's consider another example

$$\neg (A \equiv B)$$

here the negation of the biconditional is the MLF, for example, we recall that it is equivalent to number 0110 or

0110
F
T
T
F

This means that the formula is true in the second or third cases: The case where A is true and $\neg B$ is true or the cases where $\neg A$ is true and B is true. This tells us that in normal form, we can read the following equivalent statement from the truth table:

$$((A \wedge \neg B) \vee (\neg A \wedge B))$$

Again, we apply the two tree rules that we know first to the MLF (the disjunction) and then to the MLF of the two disjuncts as follows:

$$\begin{array}{ccc} & \neg(A \equiv B) & \\ & \wedge & \\ A & & \neg A \\ \neg B & & B \end{array}$$

With these examples under our belts, we can see that the following five rules suffice for tackling the job of decomposing complex formulas using the tree method.

The conjunction rule:

$$\begin{array}{c} (A \wedge B) \\ | \\ A \\ B \end{array}$$

The disjunction rule:

$$\begin{array}{ccc} & (A \vee B) & \\ & \wedge & \\ A & & B \end{array}$$

Double negation rule:

$$\begin{array}{c} \neg\neg A \\ | \\ A \end{array}$$

Rule for the negation of a conjunction:

$$\begin{array}{ccc} & \neg(A \wedge B) & \\ & \wedge & \\ \neg A & & \neg B \end{array}$$

Rule for the negation of a disjunction:

$$\begin{array}{c} \neg (A \vee B) \\ | \\ \neg A \\ \neg B \end{array}$$

14.2 How to Use the Tree Rules to Check for Validity

As we saw in previous chapters, the truth table method provides a decision procedure for checking for the validity of arguments in sentential logic, meaning that it provides a recipe, which, when followed strictly, will always allow you to determine whether a sentence is entailed by some other set of sentences. The tree method is simply an abbreviated version of the truth table method and as such can also be counted on to generate a decisive yes or no answer to the question of validity. As we shall see, the tree method, like the truth table method is automatic and requires no creativity.

The tree method proves that the conclusion follows from the premise by showing that the negation of the conclusion, together with the premises, is contradictory. If you attempt to demonstrate that your thesis must be true by showing that the negation of your thesis cannot be right, you are engaged in what is known as a *reductio* argument. The tree method is a method of indirect proof, or a *reductio*.

The tree proof works by introducing the premises and the negation of the conclusion. The negation of the conclusion is introduced as an assumption for the sake of an indirect proof. All complex formulas are then broken into their simplest constituents through the successive application of the tree rules that were introduced in Section 14.1.

Once all complex formulas have been reduced, we check each of the resulting paths for contradictions. If a path contains a contradiction then it represents an impossible way that things can be. Inspecting all the paths allows us to find states of affairs in which the premises can be true together with the denial of the conclusion. If there is no contradiction in the path, as we saw above, we call it an "open path."

If there is a way in which the premises and the denial of the conclusion can be true simultaneously, then it will show up as an open path after we apply the tree rules. If, after a complete application of the tree rules, there is an open path, then we know that the conclusion does not necessarily follow from the conclusion. The open path provides us a counterexample to the argument. The argument is invalid.

If all paths close, meaning that all paths contain some contradiction, then the argument is valid.

Let's examine an example of a proof procedure for a specific problem:

Given $(M \wedge (P \supset Q))$, $((P \vee N) \wedge \neg N)$, show Q

What is being asked here? The challenge is to show that Q follows logically from the two claims

"$(M \wedge (P \supset Q))$"

and

"$((P \vee N) \wedge \neg N)$".

STEP 1

Introduce the premises

$$\begin{gathered} M \wedge (P \supset Q) \\ (P \vee N) \wedge \neg N \end{gathered}$$

STEP 2

Introduce the negation of the conclusion, in this case "$\neg Q$," for the purpose of a *reductio* argument

$$\begin{gathered} M \wedge (P \supset Q) \\ (P \vee N) \wedge \neg N \\ \neg Q \end{gathered}$$

STEP 3

Find the MLF of one of the complex statements. Apply the tree rule for that operator to the statement as a whole. In this case, we see that the MLF of "$M \wedge (P \supset Q)$" is "$\wedge$." Following the tree rule for a conjunction, we place both conjuncts; "$P \supset Q$" and "M" in **all open paths** falling below the statement. In order to keep track of which complex statements have been tackled, it is helpful to use checkmarks. Thus, once one applies the tree rules to the MLF of a complex statement, we will mark it using a checkmark as shown below. In this case, there's only one open path and so we list M and $P \supset Q$ below the premises as follows:

$$\begin{array}{c} M \wedge (P \supset Q)\ \checkmark \\ (P \vee N) \wedge \neg N \\ \neg Q \\ M \\ P \supset Q \end{array}$$

STEP 3 (continued)

We see that "$P \supset Q$" can be resolved further. The tree rule for the conditional is to split all the open paths falling under the statement with the negation of the antecedent in one path and the consequent in the other. This is because "$P \supset Q$" is logically equivalent to "$\neg P \vee Q$." This brings us to

$$\begin{array}{c} M \wedge (P \supset Q)\ \checkmark \\ (P \vee N) \wedge \neg N \\ \neg Q \\ M \\ P \supset Q\ \checkmark \\ \diagup \quad \diagdown \\ \neg P \qquad Q \end{array}$$

STEP 4

We might notice that the path containing Q also contains $\neg Q$. This path closes, the other path remains open. To indicate a closed path, we can

$$\begin{array}{c} M \wedge (P \supset Q)\ \checkmark \\ (P \vee N) \wedge \neg N \\ \neg Q \\ M \\ P \supset Q\ \checkmark \\ \diagup \quad \diagdown \\ \neg P \qquad Q \\ \qquad\quad \text{X} \end{array}$$

STEP 5

Now we tackle "($p \vee n$) $\wedge \neg n$." Again the MLO here is the "$\wedge$" and so following the tree rule for a conjunction, we place both conjuncts; "($p \vee n$)" and "$\neg n$" in all open paths falling below the statement. In this case, there's only one open path and so we have the following:

$$\begin{array}{c}
M \wedge (P \supset Q)\ \checkmark \\
(P \vee N) \wedge \neg N\ \checkmark \\
\neg Q \\
M \\
P \supset Q\ \checkmark \\
\diagup \quad \diagdown \\
\begin{array}{cc}
\neg P & Q \\
\neg N & \text{X} \\
P \vee N &
\end{array}
\end{array}$$

STEP 6

All that's left to resolve now is "($P \vee N$)." The MLF here is the "$\vee$" and so following the tree rule for a disjunction, we place both disjuncts in all open paths falling below the statement. In this case, there's only one open path and so we have the following:

$$\begin{array}{c}
M \wedge (P \supset Q)\ \checkmark \\
(P \vee N) \wedge \neg N\ \checkmark \\
\neg Q \\
M \\
P \supset Q\ \checkmark \\
\diagup \quad \diagdown \\
\begin{array}{cc}
\neg P & Q \\
\neg N & \text{X} \\
P \vee N\ \checkmark & \\
\diagup \quad \diagdown & \\
\begin{array}{cc} P & N \\ \text{X} & \text{X} \end{array} &
\end{array}
\end{array}$$

STEP 7
Now that we have resolved all the complex statements to their simplest constituents, we examine the paths to see whether they are consistent. In both cases, the paths close because of contradictions. In the first path, we find P and $\neg P$ and in the second path, $\neg N$ and N. We had already seen that the third path closes with the contradiction between Q and $\neg Q$. Now that we have resolved all formulas and checked all paths, we can be assured of our answer. In this case, we can see that the premises of this argument and the negation of its conclusion are mutually inconsistent. Therefore, we are entitled to claim that the conclusion follows validly from the premises:

$$M \wedge (P \supset Q) \checkmark$$

$$(P \vee N) \wedge \neg N \checkmark$$

$$\neg Q$$

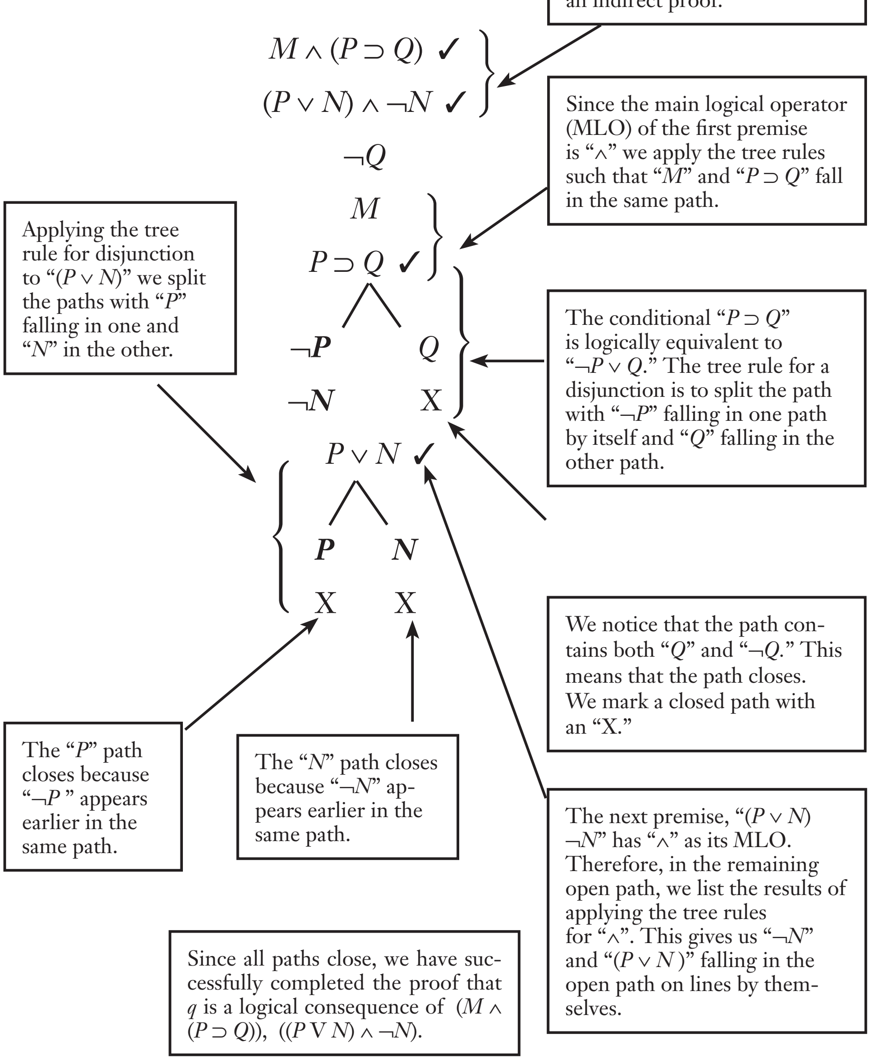
$$M$$

$$P \supset Q \checkmark$$

$$\begin{array}{cc} \neg \boldsymbol{P} & Q \\ \neg \boldsymbol{N} & \mathrm{X} \\ P \vee N \checkmark & \\ \begin{array}{cc} \boldsymbol{P} & \boldsymbol{N} \\ \mathrm{X} & \mathrm{X} \end{array} & \end{array}$$

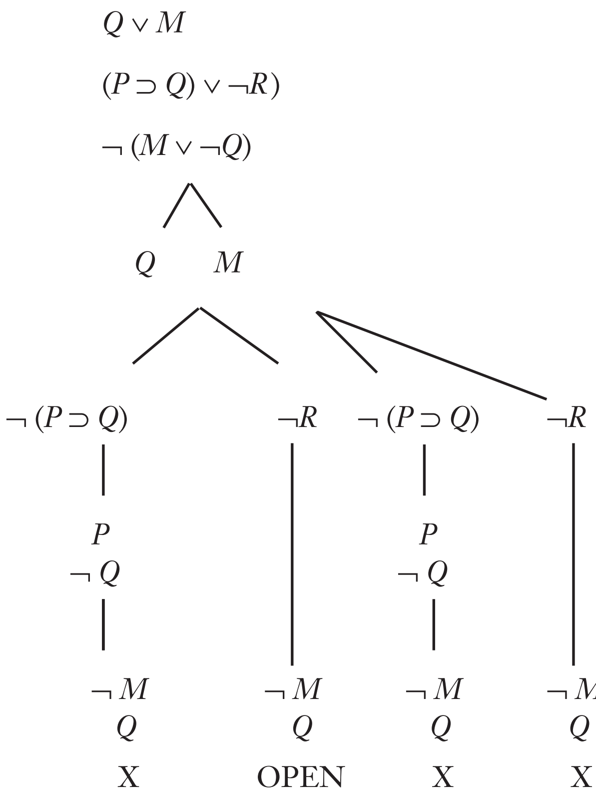
Reading a finished tree proof can sometimes require a little interpretive skill. Let's take the finished proof we've just completed:

Given $(M \wedge (P \supset Q)), ((P \vee N) \wedge \neg N)$,
Prove Q



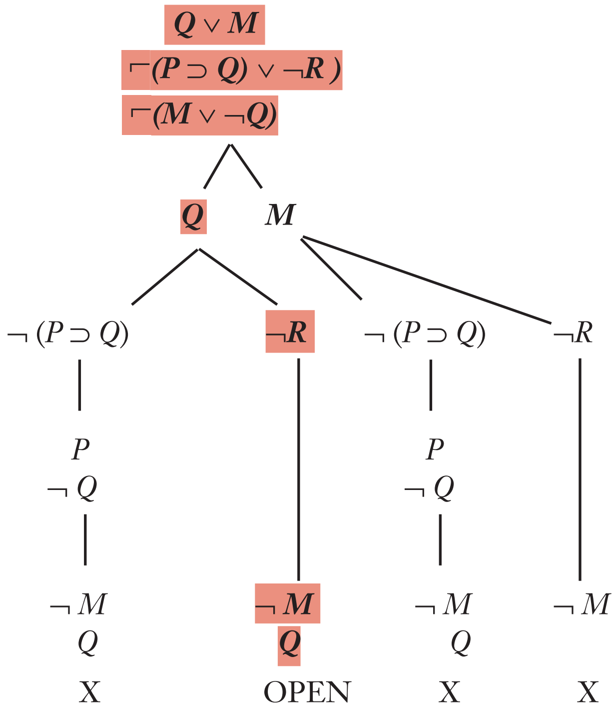
Let's consider another example:

Given $(\neg (P \supset Q) \vee \neg R)$, $(Q \vee M)$ Prove $(M \vee \neg Q)$



In this case, we have applied the tree rules to the logical operators as far as we can. We find that one of the paths remains open, meaning that it does not contain a contradiction. As we know, this means that we cannot validly conclude that $(M \vee \neg Q)$ follows from $(\neg (P \supset Q) \vee \neg R)$ and $(Q \vee M)$. The contents of the open path are the counterexample to the argument. The contents of the open path are a scenario in which the denial of the conclusion is consistent with the premises.

We can read the counterexample off from the contents of the open path. The members of this open path are highlighted below:



What we find when we read the contents of the open path is the consistent set of statements:

$$(Q \vee M), (\neg (P \supset Q) \vee \neg R), \neg (M \vee \neg Q), Q, \neg R, \neg M, Q$$

The fact that this a consistent set of sentences shows that the conclusion does not necessarily follow from the premises. Why? Because, as we can see from the open path, both $\neg M$ and Q are consistent with the premises. If that is the case, then we cannot claim that $(M \vee \neg Q)$ *must* follow from the premises.

14.3 Is There Room for Strategic Thinking in the Tree Method?

There is usually more than one way to complete a tree proof. Sometimes it will be possible to complete a tree proof with a more elegant or shorter proof depending on the order in which one's chooses to apply the rules to premises. Therefore, for example, consider the following legitimate ways of generating a proof:

Given $(\neg(\neg P \vee R) \wedge (\neg R \supset \neg(P \wedge Q))$, $((N \wedge T) \wedge \neg R)$, $(\neg Q \supset M)$
Prove $(M \wedge T)$

A.

$$(\neg(\neg P \vee R) \wedge (\neg R \supset \neg(P \wedge Q))$$
$$(N \wedge T) \wedge \neg R$$
$$\neg Q \supset M$$
$$\neg(M \wedge T)$$

$\neg M$ $\neg T$

$\neg\neg Q$ M $\neg\neg Q$ M

X (under M)

$(N \wedge T)$ $(N \wedge T)$ $(N \wedge T)$

$\neg R$ $\neg R$ $\neg R$

N N N

T T T

   X   X

$$\neg(\neg P \vee R)$$
$$(\neg R \supset \neg(P \wedge Q))$$

$\neg\neg P$

R

X

B.

$$(\neg(\neg P \vee R) \wedge (\neg R \supset \neg(P \wedge Q))$$
$$(N \wedge T) \wedge \neg R$$
$$\neg Q \supset M$$
$$\neg(M \wedge T)$$
$$|$$
$$(N \wedge T)$$
$$\neg R$$
$$|$$
$$N$$
$$T$$
$$|$$
$$\neg(\neg P \vee R)$$
$$(\neg R \supset \neg(P \wedge Q))$$
$$|$$
$$\neg\neg P$$
$$R$$
$$\text{X}$$

The difference between the first and second proof is simply the order of application of tree rules to premises. In the first example, the tree rules were applied first to $\neg(M \wedge T)$, then $\neg Q \supset M$, $(N \wedge T) \wedge \neg R$, and finally $(\neg(\neg P \vee R) \wedge (\neg R \supset \neg(P \wedge Q))$. By contrast, in B, the tree rules were applied first to $(N \wedge T) \wedge \neg R$ and then to $(\neg(\neg P \vee R) \wedge (\neg R \supset \neg(P \wedge Q))$. The second example uses a more convenient strategy simply because it involves only one path. Therefore, in general, it is a good strategy to avoid the proliferation of new paths. The simplest way to follow this strategy is to begin by trying to first apply tree rules to MLOs in a way that results in single paths rather than splitting paths too soon and dealing with all the open paths in the proof.

14.4 A Final Example

In Chapter 13, we saw examples that required the introduction of assumptions for the sake of conditional proof, such as the following:

$(A \supset (B \supset C))$

$(B \supset (C \supset D))$

$\therefore (A \supset (B \supset D))$

We saw that first one would need to assume A to derive $(B \supset C)$and then one would need to assume B in order to derive C. One needs to derive C in order to derive D. The proof ran as follows:

Show $(A \supset (B \supset D))$

1. $(A \supset (B \supset C))$	Premise
2. $(B \supset (C \supset D))$	Premise
2.1. A	Assumption CP
2.2. $B \supset C$	MP, 1 & 2.1
2.2.1. B	Assumption CP
2.2.2. $C \supset D$	MP, 2 & 2.2.1
2.2.3. C	MP, 2.2 & 2.2.1
2.2.4. D	MP, 2.2.2 & 2.2.3
2.3. $B \supset D$	CP, 2.2.1 – 2.2.4
3. $(A \supset (B \supset D))$	CP, 2.1 – 2.3

Now let's consider how this looks as a tree proof. Recall that the only assumed premise that will be introduced is the negation of the conclusion; $\neg(A \supset (B \supset D))$

$$(A \supset (B \supset C))$$

$$(B \supset (C \supset D))$$

$$\neg(A \supset (B \supset D))$$

$$A$$

$$\neg(B \supset D)$$

$$B$$

$$\neg D$$

$$(A \supset (B \supset C))$$

$$\begin{array}{cc} \neg A & B \supset C \\ \text{X} & \end{array}$$

$$\begin{array}{cc} \neg B & C \\ \text{X} & \end{array}$$

$$\begin{array}{cc} \neg B & (C \supset D) \\ \text{X} & \end{array}$$

$$\begin{array}{cc} \neg C & D \\ \text{X} & \text{X} \end{array}$$

The tree proof for this argument is elegant, simple, mechanical, and requires absolutely no ingenuity whatsoever.

15 An Introduction to First-Order Logic

15.1 Beyond Simple Declarative Sentences

It is time for us to move beyond sentential logic and into the domain of twentieth-century logic. This formalism, developed primarily by Gottlob Frege in the late nineteenth century, is called variously first-order logic, the predicate calculus, or quantification theory. The core innovation that distinguishes modern logic from its predecessors is its concern with parts of sentences rather than with whole declarative sentences. Especially, important in modern logic is its attention to quantifiers.

The quantifiers are words that indicate something about the level of generality involved in a sentence. The basic quantifiers that will serve as the focus of our study in this chapter are "some" and "all." But words like "none," "everyone," "everwhere," "nowhere," and many other terms in ordinary language also behave in ways that involve quantification. Traditional logic did not shed much insight into the logical behavior of these concepts and while Aristotelian logicians discussed reasoning involving "all" and "some" their logic was very limited in what it could formalize when compared to modern first-order logic.

This chapter introduces the needed techniques that can strengthen the expressive power of our formal reasoning. Given sentential logic alone, we will quickly encounter limits to our ability to capture some of the logical relationships and valid reasoning that we can easily express in ordinary language. While sentential logic has the wonderful property of decidability for validity it is restricted to the study of the relations between declarative sentences. However, the declarative sentence by itself is sometimes too clunky and coarse-grained for the kind of analysis we often need to do. In this chapter, we will show how to unpack the parts of a sentence in order to find the kind of logical structure that sentential logic conceals.

Remember that in sentential logic, a declarative sentence like

All frogs are amphibians.

would be represented using a single upper-case letter, say *F.* As we have seen, in sentential logic we study the logical behavior of structures that result from connecting declarative sentences using logical operators:

All frogs are amphibians or the cat ate breakfast.

Might be represented as two declarative sentences joined by the logical operator "or":

$F \vee C$

Since sentential logic deals with arguments on the scale of declarative sentences and logical operators, its resolution is limited. As we will soon see, there are a number of important features of arguments that sentential logic cannot catch.

The next step into the new world of first-order logic involves unpacking the declarative sentence in order to reveal additional its inner logical structure. As we begin to uncover the inner structure of the declarative sentence, we will find that sentences have distinct kinds of parts and that these parts interact in ways that underlie some important patterns of valid reasoning. These additional parts, the **quantifiers**, **predicates**, and **names**, add to the structure that logical operators provide. A sentence like:

<u>All</u> *frogs* are **amphibians**.

has parts that contribute to the kinds of judgments it allows us to make. Before we focus on the parts of the declarative sentence, it is worth considering what kind of logical structure sentential logic misses.

Common sense tells me that if I know that Wolfgang is a frog and I know that all frogs are amphibians, then I can validly conclude that Wolfgang is an amphibian. Notice that the word "all" is playing a crucial role in that inference. To begin with, if we are restricted to representing declarative sentences as unanalyzed units, we miss the kind of logical structure that would allow us to understand why a large class of similar arguments are valid.

Another example of the kind of logical structure that I have in mind supports the following valid judgment: Let's say I know that *Matilda bought a car*, let's call this sentence M. From M, I can legitimately conclude: *Someone bought a car*. Notice that this is a different declarative sentence, let's call it C. The inference that takes us from M to C features none of the logical operators that we have become acquainted with in the book so far. M and C are distinct, so how can I be guaranteed that this is a legitimate piece of reasoning and not a simple non-sequitur? Sentential logic alone would represent this as an inference from M to C without any additional explanation or analysis. Common sense tells us that sentential logic is missing the capacity to account for the reason we can validly move from *Matilda bought a car* to *Someone bought a car.* It seems to be a legitimate piece of reasoning by any standard, but in order to reason formally about it we will need more than sentential logic.

By the end of this chapter, you will not only be able to detect the inner structure that allows for this kind of inference, but you will also be able to formalize that structure in the language of first-order logic. In the next chapter, we will learn how to generate proofs that show the validity of arguments that depend on these structures.

15.2 Generality and Reasoning

The Matilda example above already shows how a significant limit faced by sentential logic is its inability to represent arguments involving generality. But what exactly do we mean here by generality. When we assert that all bachelors are unmarried males, we are making a general claim, but why is there any need to do more than represent this general claim as a declarative sentence? Why not simply let it be represented with a variable, say A, and then use the rules of sentential logic to study its behavior as we learned in previous chapters? As we have seen above, there is something important about the internal structure of the sentence that plays a role in reasoning that sentential logic misses, but more importantly perhaps, generality itself has logical features that make it an object of interest for the study of reasoning.

Generality has already played an important role in this book. When we say that a valid argument is one that has no counterexamples, we are making a general claim. If the premises of an argument are true and the conclusion is the result of valid inferences, then the conclusion *must* be true. Saying that it must be true, no matter what, is about as general a claim as they come.

A central feature of our reasoning is our ability to speak and think about general matters of fact. When we use words like “everyone,” “everywhere,” “everything,” “all,” and “any,” we are making claims that cast a wide net and that do not involve specific named objects, events, or places. If we say that everything is subject to the second law of thermodynamics, or that anything that has a heart has kidneys, we are stating rules or general principles that apply to all things. If we say that all humans have non-human ancestors, we are stating a rule to the effect that: If something is a human, it has non-human ancestors. Notice that we are not just talking about a long list of humans; Eric, Jessica, Sally. . . we are talking about all things. No matter who it is, if it’s a human, then it has non-human ancestors. Reasoning about general matters has its own special character. Intuitively, you already have some sense for the difference between legitimate and illegitimate ways of reasoning with generality. So for example, compare the following two pieces of reasoning:

> All students will pass the class. Therefore, since Lamar is in the class, he will pass.
>
> Bjorn has the flu. Therefore, all students have the flu.

On reflection, you can see that the first inference is valid while the second is not. In this chapter, we explore techniques that allow us to unpack and evaluate reasoning involving generality. Clearly, general claims like “all students will pass the class” can serve as the basis of valid arguments. Let’s take a slightly more complicated example:

> If I know that all whales are mammals and that at least one whale has a blowhole, it makes sense for me to conclude that at least one mammal has a blowhole.

This kind of reasoning is difficult to represent given the resources of sentential logic alone. It looks like an obviously valid argument (and it is) but it is not obvious how we might go about proving this claim given the techniques that we have studied so far in this book. Let's consider one more piece of valid reasoning using "all":

1 All tennis players admire Jones
2 One tennis player thinks Jones is arrogant
3 There is a tennis player who thinks Jones is arrogant and admires him.

Intuitively, it makes sense to say that 3 follows logically from 1 and 2. But how would you go about proving this given what we know from sentential logic? When you reflect on the reasoning from 1 and 2 to 3, it will probably occur to you that the reason this is a good inference has something to do with the role of the word "all" in this argument.

Prior to the nineteenth century, logicians struggled with what we now call **the problem of multiple generality**. This was the problem of accounting for the interplay between words like "all" and "some" in arguments that appear to be obviously valid. If, for example:

Some teacher is respected by every student.

one can legitimately conclude that:

All students respect at least one teacher.

The kind of sentential logic that we have studied so far in this book lacks the resources to explain why this is a valid piece of reasoning. In the late nineteenth century, the philosophers Gottlob Frege, a German, and Charles Sanders Peirce, an American, independently developed a way of formally representing the logic of words like "all" and "some". Pierce's contributions to logic did not get widespread attention and today Frege is widely credited as having given us a way to move beyond the limits of sentential logic and allowing us a way of understanding the logical behavior of general claims. Although his own logic is not, strictly speaking, a first-order logic, Frege's innovation gave rise to modern first-order logic in the twentieth century. Frege's most important contribution is the ability to represent the logical behavior of what came to be called the quantifiers.

Let's first introduce the symbols for the quantifiers:

"$\forall x$" reads "for all x. . ."

"$\exists x$" reads "for some x. . ."

"$\forall$" is a symbol called the **universal quantifier** and "$\exists$" is called the **existential quantifier**.

It is important not to be misled by connotations of the English word "some." When English speakers hear "some," we tend to think "more than one." However, the existential quantifier "$\exists x$" is to be understood as meaning "at least one. . . ." It is also important to recognize that when someone claims that at least one dog is friendly, they are not excluding the possibility that all dogs are friendly.

"Some" or "at least one" is not phrase that logically excludes "all."

The quantifiers will be used as parts of formal representations of sentences that talk about general features of things. Both of these symbols will be flanked on the right by one variable—either w, x, y, or z. For example, the following are all legitimate examples of quantifiers:

$\forall x$
$\exists x$
$\exists w$
$\forall y$
$\forall z$

What we mean by calling the universal quantifier "universal" is that it makes a claim universally, it applies to all things in the **universe of discourse**. As we begin to think about quantifiers, it will be necessary to be very clear about what things we think the quantifiers are referring to. For example, if I say that everyone has a mother, clearly I am not talking about cars or quasars. I am also not talking about sneezes, long walks, or the color red. When I say that everyone has a mother, you can assume from context that I am referring to a domain of discourse that includes only human beings or perhaps human beings and some animals. In some formal contexts, it will be important to clearly specify nature of the individuals we are talking about. We call this set of individuals the universe of discourse.

Of course, by itself "$\forall x$" is not a complete sentence, it simply says "for all x. . .". It needs to become part of a longer formula in order to form an assertion. Similarly, for "$\exists x$"

> Over the course of the twentieth century, different symbols have been used for the quantifiers, some authors do not use the upside down A in their text—instead of "(∀x)" they simply use "(x)"—but there is a relatively clear consensus in favor of "(∀x)" and "(∃x)" among contemporary authors and where they aren't used one can easily understand the meaning of formulas from context.

The first thing to notice about "all" and "some" is that they are interdefinable notions; they can be defined in terms of one another. In some sense, you already understand what it means to say that "all" can be defined in terms of "some" and "some" can be defined in terms of "all." If we say for instance

Not all Irish people are alcoholics.

we are saying

Some Irish people are not alcoholics.

Roughly speaking, "not all" means the same as "some are not." Similarly, if I claim

It's not the case that some bald men are atheists.

I am saying the same thing as

All bald men are not atheists.

or

No bald men are atheists.

Again, speaking roughly, "It is not the case that some. . ." means the same as "all. . . are not." We will give a more formal treatment of the interdefinability of "all" and "some" below. For now, it is enough to see that these two notions are closely related.

You have also likely noticed the role being played by the negation in these examples. Understanding the role of negation in relation to the quantifiers is critical to seeing how the logic of generality works in first-order logic. We will examine an elegant formal rule that allows us to understand why it's the case that sentences like the following are logically equivalent:

It's not the case that some high hydrants are not red.

All fire hydrants are red.

In examples like this, you can probably tell using common sense that the sentences above say the same thing. However, increasingly complicated examples with more instances of negation and quantification become difficult for common sense by itself to tackle. In order to properly articulate the logical behavior of these notions, it is first necessary to introduce some additional conventions with respect to symbolism.

Recall that in sentential logic we used variables to stand in place of declarative sentences. One major change that we will encounter as we introduce first-order logic is that instead of a single letter representing a sentence we now represent parts of our language that are smaller than complete declarative sentences. Our analysis of declarative sentences will capture the difference between the names and the properties (predicates) that figure in a sentence.

> What we mean by calling this logic "first-order" is that the things that its names and variables pick out are (at least logically speaking) *individuals*.

Recall that in sentential logic a declarative sentence like "Susan is clever" would be represented by a single uppercase letter. In first-order logic, we begin to unravel some of the inner structure of a sentence like this. So, for example, the sentence "Susan is clever" is treated in the following way: We use a lowercase letter as a constant to represent the name and an uppercase letter representing the property we are ascribing to Susan.

s will stand for "Susan"

C will stand for "clever"

The sentence "Susan is clever" will be represented as follows:

Cs

This is our first well-formed formula of first-order logic.

Notice that the letter representing the name appears to the right of the letter representing the predicate. In a moment, we will explain what's meant by "predicate" in logic. The notion of a name requires little explanation. **A name is simply a phrase that picks out an object**. In practice, naming can be fraught with complications and a quick look at the phonebook reveals that many people share the same name. We will also use names to pick out objects in addition to people. For the purposes of our study of first-order logic, we will assume that the names we use can uniquely identify objects. Our goal here is to construct a formal language that captures the logical power of ordinary reasoning, but does not suffer from its problems. In this context, we are not interested in being faithful to natural language in all its messy glory.

Names

Names are sometimes referred to as "individual constants" or "non-variable names":

We will use a lowercase italicized letter from *a* through *u* to pick out specific object.

Predicates are a little more difficult to explain. It is very common for logicians to define predicates as *properties* or *relations*. For example, in the sentence "Fido is red," the word "red" is understood to be the predicate, "Fido" is a name, and the sentence as a whole is true if and only if Fido is red. Following our convention introduced above, we will write "Fido is red" as follows:

Rf

Similarly, in the sentence "Topeka is north of Wichita," the relation "being north of" is a two-place predicate that expresses a relation between two named objects. "Topeka is north of Wichita" is true if and only if it is the case that Topeka is north of Wichita. Two-place predicates will be written as follows:

Ntw

Reading left to write, we will say that the relation *N* (is north of) obtains between *t* (Topeka) and *w* (Wichita).

A three-place predicate like "between" is involved in expressions where three named objects are involved in some relation. For example, "Albuquerque is between Denver and El Paso" is true insofar as Albuquerque actually is between Denver and El Paso.

Bade

Here, we are saying that the relationship of betweenness holds between(!) Albuqurque, Denver, and El Paso. Once we introduce four-placed predicates it becomes increasingly difficult to express the relationships in ordinary English. We will examine *n*-placed predicates in more detail below.

Predicates: In first-order logic predicates are properties of, or relations among objects. We will represent predicates using uppercase italicized letters from *A* to *Z*.

It is helpful to avoid getting too fixated on the idea of predicates representing properties and relations. As we shall see, predicates can be given an elegant formal characterization that will prove much more illuminating than the much fuzzier intuitive notions of properties and relations in the long run. The clarity of this formal account of predicates far outstrips the clarity of our grasp of notions like "property" and "relation." There are deep metaphysical questions associated with properties and relations but these can all be sidestepped for the purposes of our study of first-order logic once we have grasped the formal role of predicates.

Formally speaking, a **predicate should be understood to behave like a function that takes names as arguments and gives truth or falsity as values**.
The first step is to think about the kinds of functions you studied in mathematics. Consider the function:

$f(x) = x^2$

This function takes some number as an argument for *x* and gives as the value of the function for that argument the square of the number. If we wanted to know the value of the function for 2, for example, we would replace the *x* with 2 giving the following:

$f(2) = 2^2$

$f(2) = 4$

We know that 2^2 is 4.

As we saw in previous chapters, a function is like a machine that takes arguments as inputs and gives values as outputs. The variables serve as the slots into which we put arguments in order to get values. Predicates can have variables too. However, rather than replacing variables with numbers, in first-order logic **predicates have variables that stand in place of names for objects** and rather than having numbers as values for arguments, predicates give truth values as arguments.

> **Variables for names**: In first-order logic the lower-cased italicized letters from *v* through *z* serve as variables standing for names.

"*Rx*" is to be read as "*x* is *R*". The variable "*x*" is not a name, but it can be replaced with a name. As such "*Rx*" by itself is neither true nor false. Replacing the variable with a name, for example, "*a*" gives us "*Ra*".

Ra is either true or false depending on:

the object that *a* picks out,
the meaning of the predicate
and what the facts of the matter are

If "R" means "is red" and "a" picks out Alice the southern copperhead snake, then because it is a fact that Alice is red, "Ra" is true.

The result of replacing the "x" in "Rx" is either true or false depending on how we replace the variable "x" with a name. If Sally's Ferrari is red and we replace the "x" with the name "Sally's Ferrari", the sentence comes out true. Therefore, in this case, the predicate "is red" should be understood as a function that takes names as inputs and gives either true or false as outputs. The following table allows us to compare the kinds of things that predicates and arithmetical functions take as inputs and give as outputs.

Input (argument)	**Function**	**Output (value)**
8	$f(x) = 2x$	16
3	$f(x) = 2x$	6
Rudolph's fur	Rx	False
Rudolph's nose	Rx	True
The glass of wine I'm drinking now	Rx	True

Combining predicates, names, variables, quantifiers, and the familiar logical operators of sentential logic gives us a powerful formal language known as first-order logic. It will be important for us to express statements involving quantification as unambiguously as possible. Reviewing some of the conventions we've introduced so far:

For names or individual constants, we use lowercase letters from the beginning of the alphabet.

We reserve w, x, y, and z for variables (if we need more variables, we can add subscripts to letters, for example: w_1, w_2, w_3. . .

Uppercase letters serve as predicates.

Well-formed formulas are built up from predicates and names such that, for example:

Fa

is a well-formed formula that simply says that *a* is an *F.*

In addition to one place predicates, we can introduce *n*-place predicates standing for example for *n*-place relations. For example, in a two place predicate like "Loves" we could write the following well formed formula:

Lab

This should be read as *a* loves *b*. Any *n*-place predicate with *n* individual constants falling to its right will be considered a well-formed formula. For example,

Fa, *Gb*, *Lde*, *Rabc*, *Sdefg*

are all well-formed formulas. So if we interpret "*F*" as fantastic and *a* as Arthur, we will read "*Fa*" as "Arthur is fantastic." Once we have the atomic sentences in place, we can reintroduce the familiar logical operators of sentential logic as connectives.

In addition, we can mark the order of operations using parentheses along much the same lines as we saw in sentential logic. The purpose of introducing parentheses is to unambiguously indicate the main logical operator of the formula and to have an unambiguous hierarchy of functions. With these in place, the methods of proof that we have already mastered can be transferred into this new context.

For example, the negation of a well-formed formula is a well-formed formula

$\neg Fa$, $\neg Gb$, $\neg Lde$, $\neg Rabc$, $\neg Sdefg$

and well-formed formulas can be joined with one another using a logical operator in the same way that we joined sentence letters in sentential logic. For example:

$Rabc \supset Gb$
If *a*, *b*, *c*, are related *R-wise* then *b* is *G*

$\neg Lde \vee Mc$
d does not relate to *e L-wise* or *c* is *M*

$Pe \wedge Cp$
e is *P* and *p* is C

$(\neg Lde \vee Mc) \wedge Cp$
d does not relate to *e* *L-wise* or *c* is *M* and

etc.

It is worth slowing down a bit as we introduce the rules for building formulas that involve quantifiers and variables. Just as with sentential logic, we will be able to define what counts as a well-formed formula quite precisely.

In first-order logic, a well-formed formula (wff) is any string of symbols that obeys the following rules:

wff rule #1:
Any *n*-place predicate letter followed by *n* names is a wff

wff rule #2:
The negation of any wff is also a wff.
For example, since Gp is a wff, so $\neg Gp$

wff rule #3:
When a wff is enclosed by left and right parentheses, the resulting string is also a wff.

wff rule #4:
Given any two wffs, they can be joined by the logical operators $\wedge$, $\supset$, $\vee$, $\equiv$.

wff rule #5:

If we replace a name in a wff by a variable not previously used in the wff, while simultaneously preceding the wff with a universal quantifier for that variable, the result is a wff.

For example, because $Pf \supset Af$ is a wff, so is $(\forall x)\ (Px \supset Ax)$

Notice that we are **not** saying that from $Pf \supset Af$ we can validly derive $(\forall x)\ (Px \supset Ax)$ in this context we are just introducing the conditions that determine syntactical correctness for first-order logic; what counts as a wff, not what counts as a valid argument.

wff rule #6:

If we replace a name in a wff by a variable not previously used in the wff, while simultaneously preceding the wff with an existential quantifier for that variable, the result is a wff.

For example, because $Pf \supset Af$ is a wff, so is $(\exists x)\ (Px \supset Ax)$

wff rule #7:

Strings are only wffs if they can be built up from sentence symbols obeying wff rules 1–6 above.

Let's begin to put all the parts together in first order formulas. We begin with declarative sentences like the following:

All platonists are ambidextrous

$$(\forall x)\ (Px \supset Ax)$$

"*P*" means "platonist"
"*A*" means "ambidextrous"
The formula reads:

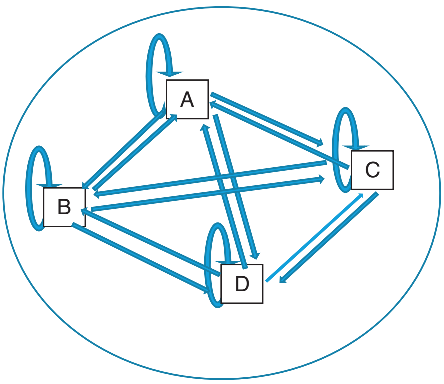
for all x, if x is a platonist then x is ambidextrous

Notice that our rules for wff permit us to have multiple quantifiers in the same formula. Notice also that we have discussed predicates that take more than one name. We used the example of "x loves y" as a way of introducing n-place predicates. As we combine multiple quantifiers and multiple variables, we will dramatically increase the expressive power of our formalism. This will prove to be the most critical advance that we find in first-order logic. Love is an intuitively accessible example. Say we want to think about the declarative sentence:

Everything loves everything.

Notice that we have a two-place predicate and we have not named anyone in particular. We are, in fact, talking about everyone and everything! This sentence is clearly false in the actual world but let us imagine conditions under which it would be true. Imagine a universe where there are only four objects. If it is true that everything loves everything, then the arrow of love travels from everything to everything else.

The world where everything loves everything:



Now that we have a model of what it would mean for it to be true that everything loves everything it is useful to think about the characteristics of that model. For example, one finds that:

(a) No matter what object one picks, there will be an arrow of love traveling from it to all other objects.

(b) No matter what object one picks, one can find an arrow of love coming to it from all other objects.

We can represent this circumstance formally in the following way:

$(\forall x)\ (\forall y)\ Lxy$

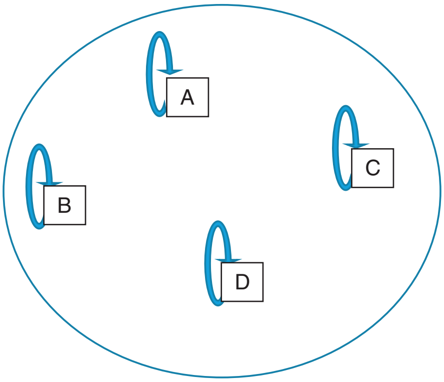
The model we presented above provides a circumstance in which the sentence is true. Logicians will say that the model *satisfies* the sentence.

We can read $(\forall x)\ (\forall y)\ Lxy$ in a variety of ways, for example, *No matter what object you pick, you can pick any object such that the relation of love obtains between your first selection and your second.*

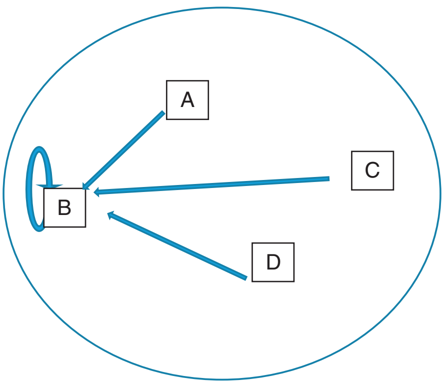
or more standardly

For all x *and for all* y, x *loves* y.

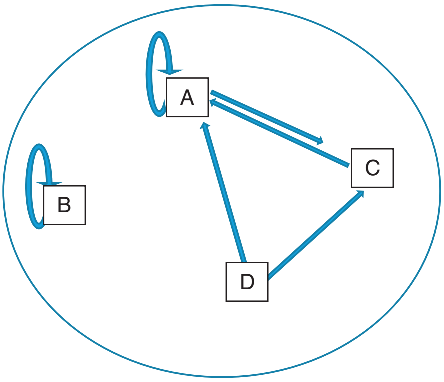
Let's consider some other way that things could be. It is certainly too optimistic to think that everything loves everything, but perhaps it is reasonable to suppose that everyone loves someone. Now notice that a variety of different kinds of worlds satisfies the claim that everyone loves someone. Consider the sad and selfish world in which everything loves only itself:



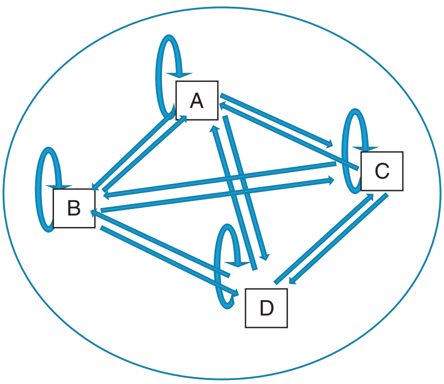
Or, the world where everyone loves one person:



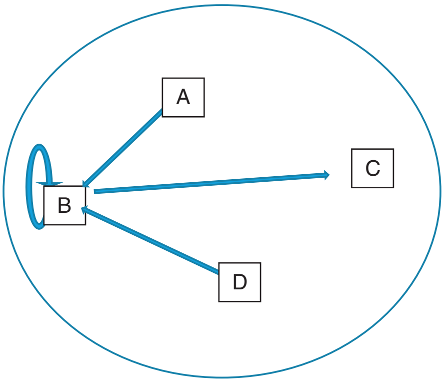
Or, the world where everyone loves someone (even though B only loves B)



Or, for that matter, the world where everyone loves everyone:



All of these models satisfy the claim that everyone loves someone. For it to be the case that the sentence is false we would have to find a model in which there are objects that do not have an arrow of love leaving them. For example, it would not be true that everyone in the following model loves someone:



In this model C is loved, but is not giving love to any objects (including itself). It would be false in this scenario to say that everyone loves someone. The way we represent "everything loves something" in first-order logic is:

$(\forall x)\ (\exists y)\ Lxy.$

One very ugly and clunky way of reading this formula is as follows: *For anything you pick, you can find something that that first thing you picked loves*. This is awkward in English because it takes so many extra words to indicate which variable we are talking about in different parts of the sentence. The formal representation, by contrast is unambiguous and elegant.

Notice the placement of the variables in this formula. Reading from the left to the right we first encounter the universal quantifier. This first quantifier is followed by the variable x. We will say that the quantifier "$(\forall x)$" **binds** the variable x; meaning that any mention of the variable x that falls in the formula that follows be understood to be the same variable that is being talked about in the quantifier. For example, in the formula that we are considering,

$(\forall x)\ (\exists y)\ Lxy.$

The x that features in the "for all x" is understood to bind the x that follows the predicate. The best way to think about binding is to recognize that the quantifiers and the variables falling after the predicates are expressing some idea together. Let's examine the way that quantifiers and predicates can express very different ideas by being arranged in different combinations. We have already seen two combinations of variables and quantifiers:

$(\forall x)\ (\exists\forall y)\ Lxy$ *For all* x *and for all* y, x *loves* y.

(Everyone loves everyone.)

$(\forall x)\ (\exists y)\ Lxy$ *For all* x, *there is at least one* y *such that* x *loves* y.

(Everyone loves someone.)

As we explore the combinations of quantifiers and variables, we will notice the expressive power and precision of first-order logic, especially when it comes to formally reasoning about relationships. Let's consider some of the other cases:

$(\exists x)\ (\exists y)\ Lxy$ *There is at least one x and at least one y such that x loves y*

(Someone loves someone.)

$(\exists x)\ (\forall y)\ Lxy$ *There is at least one x such that any y you pick, x loves it*

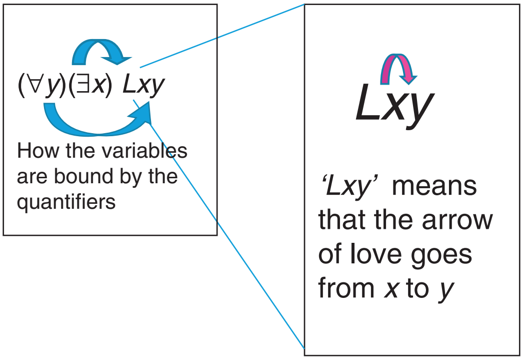
(Someone loves everyone.)

At this point, we can begin to introduce some slightly less obvious cases. Let's say we wanted to claim that everyone *is loved by* someone. Think of the kinds of worlds that would satisfy this claim: The world in which everyone is loved by their mother would make this true, as would the world in which one person loved everyone. We could represent this claim formally as

$(\forall y)\ (\exists x)\ Lxy$ *For all* x, *there is some* y *such that* y *loves* x

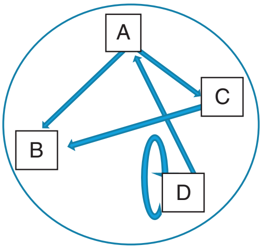
(Everyone is loved by someone.)

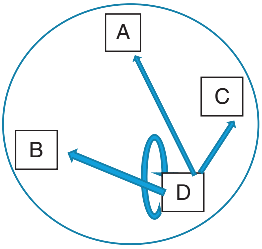
We could equally well have formalized this claim as $(\forall x)\ (\exists y)\ Lyx$. Both say the same thing. Notice that it is the interplay of the order of the quantifiers and the order of the variables following the predicate that is critical here.

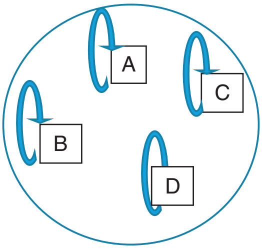


In this case, the universal quantifier binds the second variable after the predicate. The second variable is the recipient of love from the first variable following the predicate. Remember that we read "*Lxy*" as *x loves y*. As we read the formula, it is important to be aware of both the order of the quantifiers and the order in which the variables appear after the predicate.

Let's think about the wide variety of worlds that would satisfy this claim. As we saw above, the world in which everyone is loved by their mother would be one. So would the following:







It is also useful at this point to notice that any model that satisfies

$(\forall x)\ (\forall y)\ Lxy$

also satisfies:

$(\forall y)\ (\exists x)\ Lxy$

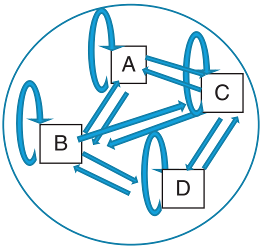
However, it is not the case that all models satisfying

$(\forall y)\ (\exists x)\ Lxy$

will satisfy

$(\forall x)\ (\forall y)\ Lxy.$

For example, the following model satisfies both, but the three previous examples satisfy $(\forall y)\ (\exists x)\ Lxy$ but not $(\forall x)\ (\forall y)\ Lxy$.



Given a two-place predicate like L, the following combinations of quantifiers and variables are wffs.

$(\forall x)\,(\forall y)\,Lxy$ Everyone loves everyone	$(\forall x)\,(\exists y)\,Lxy$ Everyone loves someone	$(\exists x)\,(\forall y)\,Lxy$ Someone loves everyone	$(\exists x)\,(\exists y)\,Lxy$ Someone loves someone
$(\forall y)\,(\forall x)\,Lyx$ Everyone loves everyone	$(\forall y)\,(\exists x)\,Lyx$ Everyone loves someone	$(\exists y)\,(\forall x)\,Lyx$ Someone loves everyone	$(\exists y)\,(\exists x)\,Lyx$ Someone loves someone
$(\forall y)\,(\forall x)\,Lxy$ Everyone is loved by everyone	$(\forall y)\,(\exists x)\,Lxy$ Everyone is loved by someone	$(\exists x)\,(\forall y)\,Lyx$ Someone is loved by everyone	$(\exists x)\,(\exists y)\,Lyx$ Someone is loved by someone
$(\forall x)\,(\forall y)\,Lyx$ Everyone is loved by everyone	$(\forall x)\,(\exists y)\,Lyx$ Everyone is loved by someone	$(\exists x)\,(\forall y)\,Lxy$ Someone is loved by everyone	$(\exists x)\,(\exists y)\,Lxy$ Someone is loved by someone

There is a lot of redundancy in this table, but it should make it clear how the interplay of quantifiers and variables works in the case of a two-place predicate. At this point, we should already be able to recognize the some of these formulas are logically equivalent and some entail the others.

15.3 The Advantages of Our Formalism Over Ordinary Language

Building a formalism from scratch has two important advantages over relying on ordinary language for thinking about statements involving generality. In ordinary language, it is often challenging (but not impossible) to eliminate ambiguity with respect to scope. A second related challenge is the occasional difficulty that we have with negation in ordinary language. Consider, for example, of the following conversation:

1. Bill: "Everyone has a chance to win the lottery."
2. Shaun: "No they don't."
3. Bill: "What do you mean? Do you think nobody is going to win?"
4. Shaun: "No, I'm just saying that they can't all win."
5. Bill: "I see what you're saying. But what if only one person played the lottery that week. Then everyone who played would win."
6. Shaun: "By 'everyone,' I thought you meant everyone, not just the people who happened to play."
7. Bill: "Oh no, I was just talking about the players."
8. Shaun: "Okay. Well I guess we agree then."

In this conversation Bill and Shaun had different interpretations of the scope of the "Everyone" in mind as it figures in line 1. Bill understood himself to be saying that each player has some chance to win the lottery. Shaun understood Bill to be saying that it is possible that everyone will win the lottery. This is the claim that Shaun meant to deny in line 2. From context, it would be reasonable to think that Shaun's interpretation of 1 is not the appropriate one, but the ambiguity in 1 means that Shaun's interpretation is permissible.

Similarly, Bill had difficulty in interpreting what Shaun meant in 3 because (without further explanation) it is ambiguous to simply deny 1 in ordinary language. Bill notices a possible interpretation of Shaun's denial which seems implausible. Can he really be denying that anyone will win? Of course, what Shaun is denying is that it is possible that everyone will win.

Eventually on line 6 and 7, they clarify the universe of discourse so that Shaun now understands Bill to have restricted consideration to the set of lottery players. With scope, negation, and universe of discourse clarified, they can see that they agree. By settling all these factors in advance, the formalism has the advantage of eliminating most of the ambiguity in ordinary language.

Let's consider an example of how scope and negation are handled in first-order logic. Consider the following sentence:

(a) Some philosophers deserve the Nobel Prize in Physics.

There are a variety of attitudes you could have to this sentence. Perhaps you agree with it. But if you said for instance that

(b) Some philosophers don't deserve the Nobel Prize in Physics.

Notice that you have not actually denied the truth of (a). (a) and (b) are compatible. The claim that some philosophers do deserve the prize is fully compatible with the fact that some do not. To deny (a) we must say:

(c) No philosopher deserves the Nobel Prize in Physics.

At this point, you are likely to be sensitive enough to scope that you are inclined to simply negate the entire sentence by adding the prefix "it is not the case that" to the entire sentence, thereby making the negation clear.

(d) It is not the case that some philosophers deserve the Nobel Prize in Physics.

In translating from ordinary language, to first-order language, it is important to be sensitive to intended meanings. As we saw in the conversation between Bill and Shaun, simply denying an existentially quantified sentence in ordinary language can be ambiguous. At this stage, we can see that responding to (a) by saying "No they don't" admits of at least the two readings (b) and (c).

Paying attention to scope becomes very important in philosophical debates concerning possibility and necessity. To see how, consider the problem of philosophical skepticism. Someone who is skeptical about the possibility of knowledge might be concerned that all of our beliefs are subject to the possibility that they are a brain in a mad scientist's vat, or that they are in the Matrix, subject to the whims of an evil God, or simply systematically defective as a knower. The skeptic might claim the following:

(1) All beliefs are possibly false.

However, notice that there is a scope issue here that alters the meaning of the claim. Does the skeptic mean to say:

(2) It's possible that *all* our beliefs are false.

meaning that all our beliefs could be simultaneously false, or the much more modest claim that

(3) Any one of our beliefs might be false.

Notice that you could deny (2) and still accept (1). (2) is a tricky claim insofar as it assumes that our beliefs do not contain a contradiction. If there is even a single contradiction in our set of beliefs, (2) cannot be true. It would be impossible for a belief and its negation to be false together! However, insofar as we think that skepticism is the view that we do not know for sure with respect to any belief taken in isolation whether it is true or false, we would be interpreting skepticism in terms closer to (3). In any event, understanding the precise extent of the skeptical worry would require the skeptic to get clear on how the quantifiers are organized in the sentence.

15.4 Negation and Quantification

In this section, we will examine a simple rule for the denial of a quantified statement. We will follow tradition in calling this rule: *Quantifier Negation* or QN for short. As we examine this rule, we will gain a clearer understanding of what it means to say that the existential and universal quantifiers are interdefinable. To begin with, consider the following:

Not all things are not squirrels.

Notice that denying that everything is not a squirrel is equivalent to claiming that there is at least one squirrel:

"$\neg(\forall x)\ \neg Sx$" is logically equivalent to "$(\exists x)\ Sx$"

Similarly, you can probably see that:

"$\neg(\exists x)\ \neg Sx$" is logically equivalent to "$(\forall x)\ Sx$"

or in English: "It is not the case that there is something that is not a squirrel" is logically equivalent to "Everything is a squirrel"

These intuitive equivalences are reflected in the formal rule of quantifier negation:

Quantifier Negation (QN):

Erase the negation symbol to the left of the quantifier, add a negation symbol to the right of the quantifier, and change the quantifier from $\forall$ to $\exists$ or from $\exists$ to $\forall$

or

Erase the negation symbol to the right of the quantifier, add a negation symbol to the left of the quantifier, and change the quantifier from $\forall$ to $\exists$ or from $\exists$ to $\forall$

For example, by the rule of quantifier negation

$\neg(\exists x)\ Sx$ becomes $(\forall x)\ \neg Sx$

$\neg(\exists x)\ \neg Sx$ becomes $(\forall x)\ \neg\neg Sx$ (which, by the rule of double negation, is logically equivalent to $(\forall x)Sx$)

$\neg\neg(\exists x)\ Sx$ becomes $\neg(\forall x)\ \neg Sx$

$\neg(\forall x)\ Sx$ becomes $(\exists x)\ \neg Sx$

etc.

QN: You can think of the rule as something like bringing the negation through the quantifier from either left to right or right to left while flipping the quantifier from ∀ to ∃ or from ∃ to ∀.

As we will see, this rule is very important as we conduct proofs in first-order logic. Notice that it also allows us a way of understanding how we can do without either ∃ or ∀. It is easy to replace any instance of $(\exists x)$ with its equivalent $\neg\ (\forall x)\ \neg$ and that we can replace $(\forall x)$ with its equivalent $\neg\ (\exists x)\ \neg$. In this book, we will use both quantifiers, but it is worth noticing that if times were tough we could economize on symbols without losing any expressive power.

15.5 Reading Formulas Containing Quantifiers

If it were true that all pirates are artists, one could conclude that anyone who is a pirate is also an artist. We have already seen the formal representation of this kind of claim in first-order logic:

$(\forall x)\ Px \supset Ax$ All pirates are artists.

Let's compare this to

$(\forall x)\ Px \vee Ax$ Everything is a pirate or an artist.

$(\forall x)\ Px \wedge Ax$ Everything is a pirate and an artist.

$(\exists x)\ Px \wedge Ax$ There is at least one pirate who is an artist.

$(\exists x)\ Px \vee Ax$ There is at least one person who is a pirate or an artist.

One way of thinking about the difference between these formulas is to compare the situations that would make them true or false. For example, think about the difference between the truth conditions for "$(\forall x)\ Px \wedge \mathrm{A}x$" and "$(\exists x)\ Px \wedge \mathrm{A}x$"

If we found just one instance of a nonpirate, then it would be false to claim that everything is a pirate. Since my sofa is not a pirate, it is not true that everything is a pirate. By contrast, in order for it to be true that there is at least one artistic pirate, we can verify the sentence simply by finding one artistic pirate.

What about the following "$(\exists x)\ Px \supset Ax$"? This is a slightly odder claim than the more familiar "$(\forall x)\ Px \supset Ax$." "$(\exists x)\ Px \supset Ax$" says something like "there exists at least one thing such that if it is a pirate, then it is artistic." This is very different from saying that all pirates are artists. Instead it is making the claim that there exists at least one thing that has the curious property of being artistic if it is piratic.

15.5.1 Negation

If we want to say that **there are no pirates**, there are two equivalent formulas in first-order logic that do the job:

$$\left.\begin{array}{l} (\forall x)\ \neg Px \\ \textit{and} \\ \neg\ (\exists x)\ Px \end{array}\right\}$$

Everything is not a pirate.
There are no pirates!
It is not the case that there is at least one pirate.

As we saw above, the rule of quantifier negation tells us that these two formulas are equivalent.

Not everything is a pirate

$$\left.\begin{array}{l} \neg(\forall x)\ Px \\ \textit{and} \\ (\exists x)\ \neg Px \end{array}\right\}$$

It's not the case that all things are pirates.
Not everything is a pirate!
Some things aren't pirates.

15.6 Instantiation

At this point, we are almost ready to begin engaging in proofs with quantifiers. The last piece of our introduction to quantifiers that we need to cover is the relationship between statements containing quantifiers and their instances. For example, if we know that

$(\forall x)\ Rx$ *Everything is ridiculous.*

Then we can replace the variable with any name while dropping the quantifier to conclude

Ra *Argle is ridiculous.*

where *a* is some name. This rule is called **Universal instantiation** or **UI**. This is simply the idea that universally quantified variable can be replaced by any name, when we say "all" we mean "all." By UI, we can validly replace a universally quantified variable with any name. By contrast, if we know that

$(\exists x)Tx$ *Something is terrifying.*

Then, given this information alone, we are *not* entitled to conclude that any named individual in particular is terrifying. We know that something is terrifying, but we are not entitled to single out Arthur, Jessica, Chaparral New Mexico, my blender, or any other named object.

However, it sometimes necessary to reason about the "someone" in question without knowing their identity and the existential quantifier does not stop us from thinking about this terrifying something without knowing precisely who or what it is.

In English, we sometimes use placeholder names like "John Doe" or "Jane Doe" when we need to talk about someone, but we do not know that person's name. In Dutch", they use "Jan Jannsen"; in French, "Jean Dupont"; and in Spanish, the phrase "Fulano de Tal." These placeholder names are ways of talking about an individual without having a name. For example, if my house is burgled, I can reason about the burglar's actions and motives while being ignorant of his or her identity. We will need a similar technique in logic. Thus, when we are instantiating an existential quantifier, we will use a **dummy name**.

In our proofs, we will sometimes need to eliminate the existential quantifiers and replace variables with dummy names. The key rule to remember when instantiating to dummy names is that we can never instantiate an existentially quantified variable to a variable or name that has been used previously in the proof. It is useful to think of existential instantiation as always being to a dummy name. However, remember that later existential instantiations should always be to new dummy names.

If I know that someone is a burglar and that someone is a philosopher who likes roses, I should be careful to instantiate the someone who is a burglar to a different dummy name than the someone who is a philosopher. Otherwise, we might illegitimately conclude that the burglar likes roses.

15.6.1 Universal instantiation

We have already seen that given a statement involving the universal quantifier as the main logical operator, we are free to drop the universal quantifier and replace the variable with any name we choose. For example, we saw that given:

$(\forall x)\ Rx$ *Everything is ridiculous.*

Then we can drop the universal quantifier and replace the variable following the predicate with any name to conclude, for example:

Ra *Argle is ridiculous.*

But what about cases where there are multiple instances of the quantified variable featuring in the formula?

$(\forall x)\ (Pxa \wedge Rx)$

Instantiating the universal quantifier means replacing all instances of the variable that fall within the scope of the quantifier. Thus, the formula above can be instantiated to the following:

$Paa \wedge Ra$

Notice that we can instantiate the variable to the name "*a*" even though "*a*" already appears in the formula. This is because when the universal quantifier is the MLF, we can instantiate the variables to *any* name.

Note that we can only apply instantiation rules for quantifiers when those quantifiers are the main logical operators in a formula. Ordinarily in first-order logic, this means the quantifiers that are located to the far left hand side of the formula over which they have scope.

Let's consider, for example:

$(\forall x)\ (\exists y)\ (Pxy \wedge Rx)$

This sentence might mean, for example, for any x you pick, you can find some y such that x pleases y and x is ridiculous. Notice that the universal quantifier binds the variable following the predicate P and the variable following the predicate R. We are talking about the same things being both ridiculous and pleasing something. Universal instantiation will apply to all instances of the bound variable in question. For example, if I wanted to universally instantiate X to the name Saul, I would apply the rules follows:

(1) $(\forall x)\ (\exists y)\ (Pxy \wedge Rx)$ Premise

(2) $(\exists y)\ (Psy \wedge Rs)$ From line 1 by Universal instantiation.

The inference from line 1 to line 2 is valid. We will begin abbreviating our justifications for inferences from here on out. For example it will suffice to justify this inference with "UI, 1". This is simply our bookkeeping method for keeping track of how we got each line in the proof. Notice how the outermost quantifier drops out and the variables that are bound by (fall within the scope of) the quantifier are replaced by a name. We must replace all the bound, universally quantified variables falling within the scope of the universal quantifier consistently. It would not be permitted to replace the first x after the predicate P with a different name than the x appearing after the predicate R. Note that nothing stops us from repeated uses of the rule of Universal instantiation on line 1.

(1) $(\forall x)\ (\exists y)\ (Pxy \wedge Rx)$	Premise
(2) $(\exists y)\ (Psy \wedge Rs)$	UI, 1
(3) $(\exists y)\ (Pay \wedge Ra)$	UI, 1
(4) $(\exists y)\ (Pby \wedge Rb)$	UI, 1

We can continue generating instances by the rule of Universal instantiation from line 1 to our hearts content. But what is the purpose of engaging in Universal instantiation? Think about cases where we are reasoning about general matters. For example, take our earlier sentence:

All frogs are amphibians.

If we know that Stanley is a frog, then we can conclude that he is an amphibian. How do we represent this kind of reasoning in first-order logic? As we saw above, we can represent it in the following way:

$(\forall x)\ Fx \supset \mathrm{A}x$

This reads, for all x, if x is a frog, then x is an amphibian.
Now given that we know that Stanley is a frog.

Fs

How does one arrive at the intuitive conclusion

As

that Stanley is an amphibian? Proofs of this sort run along the following lines:

(1)	$(\forall x)\ Fx \supset Ax$	Premise
(2)	Fs	Premise
(3)	$Fs \supset As$	UI, 1
(4)	As	MP, 2,3

Informally what this says is that given the two premises; (1) all frogs are amphibians and (2) Stanley is a frog, we can infer by Universal instantiation of the variable x in the first premise that (3) if Stanley is a frog, then Stanley is an amphibian. Finally, given lines 2 and 3, we can conclude by modus ponens that (4).

15.6.2 Existential instantiation

If our proof requires us to consider an instance of an existentially quantified variable, as we saw above, we must instantiate the variable to a dummy name. For example, consider the following statement:

$(\exists x)\ (Pbx \wedge Rx)$

This can be instantiated to a dummy name, say for example, *d*. Giving us:

$Pbd \wedge Rd$

It is important to keep in mind that by using the dummy name instead of the existentially quantified variable, we are not naming any known object or person. Thus, the restriction on existential instantiation is simply that one cannot use a name that has appeared previously in the proof. One should not existentially instantiate to a name that appears in the premises, nor can one existentially instantiate to a name that is introduced as part of an assumed premise.

It is easy to understand why we need this restriction. Think of the following set of premises:

Someone stole my bike and broke my gate. $(\exists x)\,(Sx \wedge Bx)$

Anyone who steals a bike is probably desperate. $(\forall x)\,(Sx \supset Dx)$

Frank was seen outside my house and is desperate. $Ox \wedge Dx$

There is no legitimate way to infer from the fact that someone stole my bike and broke my gate; the conclusion is that Frank stole my bike. However, I can conclude that the person who broke my gate was probably desperate.

1. $(\exists x)\,(Sx \wedge Bx)$	Premise
2. $(\forall x)\,(Sx \supset Dx)$	Premise
3. $Sd \wedge Bd$	EI, 1 (d)
4. Sd	S, 3
5. $Sd \supset Dd$	UI, 2
6. Dd	MP, 4 & 5
7. $Sd \wedge Dd$	Conj 4 & 6

In practice, we can be guided by a strategic principle that says that **we should perform our existential instantiations prior to our universal instantiations**. At this point, we can use those methods for derivation of conclusions that we learned earlier, direct, conditional, or reductio-style proofs for first-order logic.

15.6.3 Universal generalization

The additional rules of inference that we have introduced, Universal Instantiation and Existential Instantiation can be supplemented with two more rules; **Universal Generalization** and **Existential Generalization**. The first of these, universal generalization, involves moving from arbitrary names to a general rule. This can look like an instance of fallaciously hasty generalization, but under certain restricted conditions we can move from

Sa to $(\forall x)\,(Sx)$

$$\frac{Sa}{(\forall x)\,(Sx)}$$

It is worth repeating that this can only happen under very restricted circumstances: Provided that *a* is an individual that has not figured in the premises of the argument nor has it appeared previously in a subproof of a conditional proof or an indirect proof. The key idea here is that we can only move from a statement about some individual to a generalization if that individual is an *arbitrary individual*, meaning that it is not used to refer to anyone in particular. This is a little tricky and it is easy to see how one could go wrong using the rule of universal generalization, but the rationale for it is straightforward. Consider the following valid inference:

All whales are mammals. $(\forall x)\ (Wx \supset Mx)$

All mammals have hearts. $(\forall x)\ (Mx \supset Hx)$

Therefore, all whales have hearts. $(\forall x)\ (Wx \supset Hx)$

Without the rule of universal generalization, it would be difficult to derive this conclusion using the rules discussed above. Let's formalize it in order to see where universal generalization would play a role. Once you see how universal generalization works and where it is needed, you can more easily spot illegitimate instances of generalization that violate the rule.

1. $(\forall x)\ (Wx \supset Mx)$	Premise
2. $(\forall x)\ (Mx \supset Hx)$	Premise
3. $Wa \supset Ma$	UI, 1
4. $Ma \supset Ha$	UI, 2
5. $Wa \supset Ha$	HS, 3 & 4
6. $(\forall x)\ (Wx \supset Hx)$	UG, 5

The reason that it is not problematic to apply the rule of universal generalization from line 5 in line 6 is due to the fact that *a* plays the role of an arbitrary individual introduced by universal instantiation from lines 1 and 2. The way to think of an arbitrary individual here is to imagine the following line of reasoning:

All whales are mammals and all mammals have hearts. Let's say there's someone named Abra, if Abra is a whale, then Abra is a mammal and if Abra is a mammal, then Abra has a heart. This means that if Abra

> *is a whale, then Abra has a heart. Abra is an arbitrary individual, what is true of Abra is true of anyone. Therefore, if anyone is a whale, then they would have a heart.*

Universal generalization is sometimes described as characterizing mathematical reasoning in which, for example, geometrical proofs involving a specific arbitrary case are taken to license general claims about all such cases.

However, the rule of universal generalization should be treated with great care. When using it in a proof, it should be completely clear that one is universally generalizing from an arbitrary individual and that the name in question has not played a role in the proof that would generalize illicitly from properties of particular named individuals to all individuals. For example, we cannot reason from the fact that Chaz the whale is a mammal and has a heart to the conclusion that all whales have hearts. Notice how the name, Abra, above appears as the result of a universal instantiation. There are no facts about Abra that wouldn't also be true of anyone else. In this sense "Abra" names an arbitrary individual whereas "Chaz" names a whale. We know nothing specifically about Abra, nor do we *need* to know anything about Abra in order for us to be able to say that: *If Abra were a whale, then Abra would be a mammal.*

15.6.4 Existential generalization

Existential generalization is, thankfully, a far more straightforward rule than universal generalization. It is simply the valid inference from the fact that some named individual has some property to the claim that something has that property. For example, from the fact that Chaz is a whale, I can validly conclude that something is a whale

$$\frac{Wc}{(\exists x)\,(Wx)}$$

15.7 Tree Proofs in First-Order Logic

Once rules for universal and existential instantiation and quantifier negation are added to our repertoire and we have some experience with derivations in first-order logic, we can reintroduce the tree rules and can use them to construct proofs in first-order logic.

As with the tree rules for sentential logic, we can construct a reductio proof for some conclusion by demonstrating that the premises and the negation of the conclusion are inconsistent. Once again, the goal of our tree proof is to show that there is no way that the premises and the negation of the conclusion can be true together. We will attempt to close all paths in the tree, thereby demonstrating the validity of the argument.

Let's consider an example: Given $(\forall x)\ (\neg Rx \wedge Qx)$ and $(\forall x)\ (\forall y)\ (Pxy \vee Rx)$ let's say that we need to prove that *Pab* follows validly. Once again we will assume the negation of the conclusion to test for counterexamples to the argument, then we will apply the tree rules to the MLF of the formulas successively until there are no more functions to deal with.

With the introduction of quantifiers, we must remember to treat them as MLF where appropriate. In this example, the MLF of both premises is a universal quantifier. This means that in both cases, we will need to apply the rule of universal instantiation before we can do anything else to the premises.

$(\forall x)\ (\neg Rx \wedge Qx)$	Premise
$(\forall x)\ (\forall y)\ (Pxy \vee Rx)$	Premise
$\neg Pab$	Denial of the conclusion
$\neg Ra \wedge Qa$	UI from $(\forall x)\ (\neg Rx \cdot Qx)$
$\neg Ra$	Simplification of $\neg Ra \wedge Qa$
Qa	Simplification of $\neg Ra \wedge Qa$
$(\forall y)\ (Pay \vee Ra)$	UI from $(\forall x)\ (\forall y)\ (Pxy \vee Rx)$
$Pab \vee Ra$	UI from $(\forall y)\ (Pay \vee Ra)$
/ \	Paths split because the MLF of $Pab \vee Ra$ is a disjunction
$PabRa$	Paths close (argument is valid)
x x	

This is a relatively straightforward proof, but now let's consider, what happens in cases where there are formulas with multiple quantifiers. The following is an invalid argument. Let's see, how we might try to demonstrate the existence (or lack thereof) of a counterargument.

$(\exists x)\ (\forall y)\ (Lxy \wedge Ray),\ (\forall x)\ (\forall y)\ (Lxy \vee Fxy),\ \neg Fcg \ \therefore Rab \wedge Lde$

$$(\exists x)\ (\forall y)\ (Lxy \wedge Ray)$$
$$(\forall x)\ (\forall y)\ (Lxy \vee Fxy)$$
$$\neg Fcg$$
$$\neg(Rab \wedge Lde)$$
$$(\forall y)\ (Ldy \wedge Ray)$$
$$Lde \wedge Rae$$
$$Lde$$
$$Rae$$
$$(\forall y)\ (Ldy \vee Fdy)$$
$$Lde \vee Fcg$$

Lde	*Fcg*
	X

$\neg Rab$	$\neg Lde$
OPEN	*X*

Again, we followed the strategic rule that tells us to perform our existential instantiations prior to performing universal instantiations. While we demonstrated the existence of a counterargument here, the existence of open paths has a somewhat different significance in the case of first-order logic. Unlike the case of sentential logic, **first-order logic requires a little ingenuity and is not decidable**. We do not have a mechanical procedure for checking for validity in first-order logic that will work for all cases. In practice, this means that an open path may have resulted from poor strategic choices earlier in the proof. It might be the case that an open path continues without end because the correct strategy for generating a contradiction was not spotted at an earlier stage.

First-order logic is sometimes called polyadic quantification theory in order emphasize the importance of its ability to capture the logical interplay among sequences of nested quantifiers. In the proof that follows, we see how multiple instantiations can be played out via the tree method.

$(\exists x)\ (\exists y)\ (\forall z)\ ((Mxyz \supset Hx) \wedge (Rz \wedge Mxyz)),\ (\forall x)\ (Hx \supset \neg Pc)\ \therefore \neg Pc$

Tree	
$(\exists x)\ (\exists y)\ (\forall z)\ ((Mxyz \supset Hx) \wedge (Rz \wedge Mxyz))$	Premise
$(\forall x)\ (Hx \supset \neg Pc)$	Premise
$\neg\neg Pc$	Denial of the conclusion
$(\exists y)\ (\forall z)\ ((Mdyz \supset Hd) \wedge (Rz \wedge Mdyz))$	EI from the first premise
$(\forall z)\ ((Mdd_1z \supset Hd) \wedge (Rz \wedge Mdd_1z))$	EI to another dummy name: d_1
$((Mdd_1a \supset Hd) \wedge (Rz \wedge Mdd_1a))$	UI from the previous line
$Mdd_1a \supset Hd$	Application of the tree rule for $\wedge$
$Rz \wedge Mdd_1a$	Application of the tree rule for $\wedge$
Rz	Application of the tree rule for $\wedge$
Mdd_1a	Application of the tree rule for $\wedge$
	Paths split because of application of tree rule for $\supset$ to $(Mdd_1a \supset Hd)$
$\neg Mdd_1a$ X Hd	First path closes, second is open
$Hd \supset \neg Pc$	UI from the second premise
$\neg Hd$ $\neg Pc$	Application of the tree rule for $\supset$
X X	Both paths close

The tree method offers a great deal of formal power in the construction of proofs. While they sometimes appear cumbersome on the printed page, they offer a relatively simple procedure that exploits the power of the reductio argument. By way of example, we can revisit our discussion of the rule of universal generalization above. When the rules of universal and existential generalization were introduced above, readers might have noticed that no examples were used to illustrate the rules in action. Instead, the rules were introduced, explained, and justified. This is simply because once we have the tree method in place, we can do without the rule of universal generalization. Let's think about the kind of example that is often used to motivate the introduction of universal generalization: Recall our example above: All whales are mammals $(\forall x)\ (Wx \supset Mx)$, All mammals have hearts $(\forall x)\ (Mx \supset Hx)$, Therefore, all whales have hearts $(\forall x)\ (Wx \supset Hx)$. We can demonstrate the validity of this argument very straightforwardly using the tree method:

$(\forall x)\ (Wx \supset Mx),\ (\forall x)\ (Mx \supset Hx)\ \therefore\ (\forall x)\ (Wx \supset Hx)$

$$
\begin{array}{c}
(\forall x)\ (Wx \supset Mx) \\
(\forall x)\ (Mx \supset Hx) \\
\neg\ (\forall x)\ (Wx \supset Hx) \\
(\exists x)\ \neg(Wx \supset Hx) \\
\neg(Wd \supset Hd) \\
Wd \\
\neg Hd \\
(Wd \supset Md) \\
\begin{array}{cc}
\diagup & \diagdown \\
\begin{array}{c} \neg Wd \\ \text{X} \end{array} &
\begin{array}{c} Md \\ Md \supset Hd \\
\begin{array}{cc}
\diagup & \diagdown \\
\begin{array}{c} \neg Md \\ \text{X} \end{array} & \begin{array}{c} Hd \\ \text{X} \end{array}
\end{array}
\end{array}
\end{array}
\end{array}
$$

This concludes our whirlwind tour of first-order logic. The purpose of this introduction to the logic of the quantifiers was to demonstrate the power of a formal method that goes beyond sentential logic in sophisticated and interesting ways. It should also be clear that the representational apparatus of first-order logic far outstrips the power of ordinary languages like English to represent complex patterns of dependence and independence between quantifiers. It is difficult to construct English sentences that express four or more place predicates, for example. In first-order logic this is a simple matter. Additionally, patterns of dependence and independence between these four quantifiers can be expressed in ways we explored in the much simpler context of love (our favorite two-place predicate) above.

Bibliography

Benartzi, Shlomo, and Richard H. Thaler. "Myopic loss aversion and the equity premium puzzle." The Quarterly Journal of Economics 110.1 (1995): 73–92.

Brouwer, L. E. (1907). On the foundation of mathematics. Collected Works, 1, 11–101.

Chaffee, John. "Teaching critical thinking across the curriculum." New Directions for Community Colleges 1992.77 (1992): 25–35.

DeMay, Richard M. "Should we abandon pap smear testing?" American Journal of Clinical Pathology. Pathology Patterns Reviews.114. Suppl 1 (2000): S48–S51.

Enderton, Herbert, and Herbert B. Enderton. *A mathematical introduction to logic*. Academic press, 2001.

Griffin, L. Y., Agel, J., Albohm, M. J., Arendt, E. A., Dick, R. W., Garrett, W. E., & Wojtys, E. M. (2000). Noncontact anterior cruciate ligament injuries: risk factors and prevention strategies. Journal of the American Academy of Orthopaedic Surgeons, 8(3), 141–150.

Hintikka, Jaakko, and James Bachman. *What If--?: Toward Excellence in Reasoning*. McGraw-Hill Humanities, Social Sciences & World Languages, 1991.

Kahneman, Daniel, and Amos Tversky. "Subjective probability: A judgment of representativeness." The concept of probability in psychological experiments. Springer Netherlands, 1972. 25–48.

Kahneman, Daniel, and Amos Tversky. "Prospect theory: An analysis of decision under risk." Econometrica: Journal of the econometric society (1979): 263–291.

Kunda, Ziva; Fong, G.T.; Sanitoso, R.; Reber, E. (1993), "Directional questions direct self-conceptions". Journal of Experimental Social Psychology (Society of Experimental Social Psychology) 29: 62–63.

Lord, Charles G., Lee Ross, and Mark R. Lepper. "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence." Journal of Personality and Social Psychology 37, 11 (1979): 2098.

Mercier, Hugo and Dan Sperber. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory (June 26, 2010)." Behavioral and Brain Sciences 34(2), 57–74. Available at SSRN: *http://ssrn.com/abstract=1698090*.

Miller, George A. "The magical number seven, plus or minus two: some limits on our capacity for processing information." Psychological Review 63.2 (1956): 81.

Nickerson, Raymond S. "Confirmation bias: A ubiquitous phenomenon in many guises." Review of General Psychology 2.2 (1998): 175.

Novaes,Catarina Dutilh. Formal languages in logic: A philosophical and cognitive analysis. Cambridge University Press, 2012.

Quine, Willard Van Orman. *Methods of logic*. Harvard University Press, 1982.

Robinson, Dwight E. "Fashions in shaving and trimming of the beard: The men of the illustrated London news, 1842–1972." American Journal of Sociology (1976): 1133–1141.

Simon, Herbert Alexander. Models of bounded rationality: Empirically grounded economic reason. Vol. 3. MIT press, 1982.

Walton, Douglas, Christopher Reed, and Fabrizio Macagno. Argumentation schemes. Cambridge University Press, 2008.

Wason, P. C. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20(3):273–281.

Wason, Peter C. “On the failure to eliminate hypotheses in a conceptual task.” Quarterly Journal of Experimental Psychology 12.3 (1960): 129–140.

Willingham, Daniel T. “Critical thinking: Why is it so hard to teach?.” Arts Education Policy Review 109.4 (2008): 21–32.

Wilson, Robert C., et al. “Humans use directed and random exploration to solve the explore–exploit dilemma.” Journal of Experimental Psychology: General 143.6 (2014): 2074.