forallχ

Leeds PHIL1250 2022-23

P.D. Magnus

University at Albany, State University of New York

Modified for the University of Leeds PHIL1250 module by:

E.J.R. Elliott

University of Leeds

P. D. Magnus would like to thank the people who made this project possible. Notable among these are Cristyn Magnus, who read many early drafts; Aaron Schiller, who was an early adopter and provided considerable, helpful feedback; and Bin Kang, Craig Erb, Nathan Carter, Wes McMichael, Selva Samuel, Dave Krueger, Brandon Lee, Toan Tran, and the students of Introduction to Logic, who detected various errors in previous versions of the book.

E. J. R. Elliott would like to thank P.D. Magnus for his generosity in making forallχ available to everyone. He would also like to thank Jessica Isserow for discussion and helpful comments on drafts.

The style for typesetting proofs is based on fitch.sty (v0.4) by Peter Selinger, University of Ottawa.

How to Read This Book

There are eight main chapters, corresponding to eight lectures. Together these are designed to provide you with the kind of background in formal logic and probability theory that will be useful for more advanced study in philosophy. They are not designed to constitute a complete course in logic or probability theory. We have focused primarily on the basic concepts and formalisation.

For those wishing to delve into proof systems, you can find two supplemental chapters on these in the appendices. You are not required to read these chapters, and you will not be examined on anything contained therein. You can also enrol in the formal logic module for next year, where you'll learn about many of the things discussed in this book in much greater detail.

Important concepts are often **bolded** when they're first introduced. Furthermore, throughout the chapters key ideas will often be summarised in boxes, like so:

> A key idea will often be summarised here. You are advised to pay close attention to what's written inside a box like this one.

Every chapter also ends with some practice problems. The answers to some but not all of these problems are provided in an appendix. The problems that have solutions in the appendix are marked with a star '$\star$'. In some cases, more than one solution to a practice problem might be possible; where this is the case, the solutions in the appendix only represent one possible way of proceeding.

Finally, it is the nature of logic texts that they may sometimes contain errors. If you happen to spot what you think is an error, then please do let me know.

Contents

Part I

Logic

Chapter 1

What is Logic?

Central to the study of logic is the evaluation of arguments. In everyday language, we sometimes use the word 'argument' to refer to heated disagreements between two or more people. If you and a friend have an argument in this sense, then things are perhaps not going so well between the two of you. However, in the study of logic, we are not interested in the teeth-gnashing, anger and insults kind of argument. Rather, we are interested in logical arguments.

This introductory chapter starts with a review of some of the most important concepts involved in the evaluation of arguments. We begin the study of formal logic with a clear understanding of what arguments are, of what it means for an argument to be sound and valid, and of what it is for arguments to have a valid formal structure. Later, we will start to translate arguments from English into a simple formal language.

1. Arguments, Sentences, and Propositions

Here is an example of a very simple argument:

P1 No snakes have fur.
P2 All pythons are snakes.

C No pythons have fur.

Here, P1 and P2 are the **premises** of the argument, and C is the **conclusion**. The line between the premises and the conclusion is called an **inference bar**; it can be taken to indicate that the premises above it are supposed to support—roughly: give you reasons for accepting—the conclusion that follows below it. In this case, you should get an intuitive sense as to how P1 and P2 jointly support the conclusion C.

In this text we will define an **argument** as an ordered sequence of two or more propositions, exactly one of which is the conclusion and the rest of which are premises. The final proposition in the sequence is always the conclusion. (I'll explain in a moment what a proposition is.) This is the definition we'll be using throughout the text. There are a couple of things interesting things that you should note about the definition:

- ▹ An argument always has exactly one conclusion. Of course, the conclusion of one argument might also sometimes serve as a premise in some other argument; because of this, it is possible to chain arguments together into an extended argument. Throughout this text, though, we will restrict our attention to formalising arguments that have one and only one conclusion.
- ▹ Some logic textbooks define 'argument' such that there can be zero premise arguments. It is a little strange to talk about arguments with no premises at all, but for reasons of convenience and generality logicians will sometimes want to say that there can be zero-premise arguments. We use a different definition of 'argument' that requires each argument to have at least one premise. There is no fact of the matter as to which definition is correct, just different conventions.

The definition of an argument makes reference to propositions—so what's a **proposition**? That's a tricky question. The metaphysics of propositions is a matter of contemporary debate among philosophers. For our purposes, it will be enough if we say that a proposition is what is expressed by a declarative sentence as uttered on a given occasion.

Ok—so what's a **declarative sentence**? It is the kind of sentence that can sensibly be said to be either true or false. In the logician's jargon, declarative sentences are said to be **truth-apt**. This doesn't mean that the sentence *is* true, just that it's the kind of sentence that could sensibly be either true or false. (It would be just as good to say that declarative sentences are falsity-apt, since any sentence that is truth-apt must also be falsity-apt.) For example, 'The moon is round' is a declarative sentence. It expresses the proposition *the moon is round*. In other words, the proposition *the moon is round* is the meaning of the sentence, 'The moon is round'. The sentence happens to be true because things really are as the sentence says they are; if the moon were a pyramid then the sentence would be false. Other declarative sentences include:

- ▹ 'The moon is made of cheese.'
- ▹ 'An asteroid wiped out the dinosaurs.'
- ▹ 'Dogs and cats are not usually purple.'
- ▹ 'All humans are either mortal or immortal.'
- ▹ 'If I am well-fed, then I am happy.'

Each of these express a different proposition, some true and some false. Contrast these with other types of sentences in English that are non-declarative:

INTERROGATIVES: Questions such as 'Are you sleepy yet?' or 'Have you done your logic homework?' are examples of interrogative sentences. It does not make sense to say that a question either true or false—the *answer* to a question may be either true or false, but the *question* itself cannot be. So interrogative sentences are not truth-apt.

IMPERATIVES: Commands such as 'Wake up!', 'Sit up straight!', or 'Do your logic homework!' are examples of imperative sentences. Although it might be a good thing for you to do your logic homework, it does not make sense to say that the command itself is either true or false. So imperative sentences are not truth-apt.

EXCLAMATIONS: Expressions such as 'Ouch!' and 'Boo!' are examples of exclamatory sentences. If someone just yelled out 'Ouch!', it would not make sense to respond 'Yes that's true'. So exclamatory sentences are not truth-apt.

Interrogative, imperative, and exclamatory sentences are all non-declarative—they do not express propositions, and so they are not fit to express the premises and/or the conclusions of any logical argument. Hence, the simple formal languages that we will develop later on in this book will not deal with them; we focus on declarative sentences only.

It is important to make sure you understand the difference between *sentences* and *propositions*. A single proposition can be expressed by many different sentences. For example, the English sentence 'Snow is white' expresses the very same proposition as the English sentence 'White is the colour of snow', which expresses the same proposition as the French sentence 'La neige est blanche'. All of them express the proposition that *snow is white*. Furthermore, some special sentences—for example, those involving indexical and demonstrative terms like 'I', 'you', 'here', 'now', 'that', 'there'—can express different propositions depending on the context in which they are uttered. For example, if Frank were to say 'I am happy' then he would express the proposition that *Frank is happy*, whereas if Jackson were to say 'I am happy' then he would express the proposition that *Jackson is happy*.

So, while propositions and declarative sentences are very closely connected to one another, they are not the same thing and you should be careful not to conflate them. Some propositions are expressed by many different sentences, and some sentences can be used (in different contexts) to express many different propositions.

Above, we defined an argument as a sequence of propositions. Since propositions can expressed by declarative sentences, it follows that arguments can be expressed by sequences of declarative sentences. All of the ordinary language arguments you'll find in this textbook is expressed using a sequence of sentences of ordinary English. Consequently, in the chapters that follow we will spend a lot of time talking about the meanings of ordinary English sentences. The study of logic is very closely connected to the study of meaning and languages. Indeed, many of the ideas we'll be discussing below are ideas you already understand (in one form or another, perhaps implicitly) by virtue of being a competent speaker of English.

> A PROPOSITION is the meaning of a declarative sentence, as uttered on a given occasion. A DECLARATIVE SENTENCE is a truth-apt sentence; that is, it is the kind of sentence that can be either true or false.

We say that *true* (T) and *false* (F) are **truth-values**. In the kinds of logical systems that we will be considering in this text, we assume that every proposition—and therefore every declarative sentence—must be either true or false. Nothing can be both true and false, nothing can be neither true nor false, and there is no in-between.

(Classical logics are built on these assumptions. There are also many varieties of 'non-classical' logics that, for example, have more than two truth-values or allow some sentences and propositions to have more than one truth-value or perhaps no truth-value at all. Some logics posit three truth-values, some have four or five, and some have infinitely many truth-values. But we need to walk before we can run, and the first step to learning any non-classical logic is becoming familiar with the classical two-valued system.)

> There are two TRUTH-VALUES. In classical logic, a proposition can be either true or false; furthermore, a proposition cannot be both truth and false, and it cannot be neither true nor false.

2. Validity and Soundness

Our definition of an argument is very general. Consider the following:

P1 There is coffee in the coffee pot.
P2 There is a dragon playing bassoon in the attic.

C Dali is a poker player.

It may seem odd to call this an argument. Certainly, you wouldn't find many people willing to say that the premises P1 and P2 give us good reasons to believe the conclusion. However, given our definition, it counts as an argument. It is a sequence of two or more propositions (expressed by declarative sentences), of which exactly one is a conclusion and the rest of which are premises. The following also counts as an argument:

P1 There are dogs.
P2 If there are dogs then there are animals.
P3 If there are animals then there are things.

C There are things.

This argument is much better than the first one. But what does this mean—what is it for an argument to be a *good* argument? We can start to answer that question by considering what goes wrong with the following argument:

P1 It is raining outside.
P2 If you do not have an umbrella, then you will get wet.

C You must wear sunglasses.

Suppose that P1 is true, and it really is raining outside. Is P2 true? Not clearly. Instead of using an umbrella to stay dry, you might instead try using a raincoat, or you might just stay under cover away from the rain. Using an umbrella is not the only way to stay dry when it's raining, so P2 is too strong. Suppose, however, that both of the premises really are true. It really is raining outside, and you will indeed get wet if you do not have an umbrella. (For example, if you are already outside away from cover, and you do not have a raincoat.) Does the conclusion follow? Clearly not.

The example suggests that there are at least two kinds of problems an argument might have—two distinct ways in which an argument might be considered flawed or inadequate or otherwise less-than-ideal: (1) one or more of its premises might be false, and (2) the premises may fail to guarantee the conclusion, *regardless of whether those premises are true.*

Consider another example:

P1 Frank is reading a logic textbook.
P2 Everyone who reads logic textbooks is a logic student.

C Frank a logic student.

This argument is (deductively) **valid**. An argument is valid in this sense when the truth of its conclusion is logically guaranteed by the truth of its premises. To say the same thing in a different way: an argument is valid just in case it is impossible for all its premises to be true and its conclusion to be false. Every argument is either valid or invalid; it must be one or the other, and it cannot be both.

When judging validity, you don't need to worry about whether the premises are *actually* true—what matters is whether those premises guarantee the conclusion on the assumption of their truth. This means that arguments can sometimes have true premises and a true conclusion, and yet still be invalid. For instance:

P1 London is in England.
P2 Beijing is in China.

C Dogs exist.

Even though both of the premises are true, and the conclusion is also true, the argument is clearly invalid because it is *possible* for the premises P1 and P2 to both be true even while the conclusion C is false. The important point about a valid argument is that *if* all of the premises are true, then the conclusion *must* also be true. Whether the premises *are in fact* true does not make a difference to whether an argument is valid or invalid.

A valid argument can also sometimes have a false conclusion. For example, the following argument is valid, but has a false conclusion:

P1 Dogs do not exist.
P2 Either dogs do exist or Aristotle a famous guitarist.

C Aristotle was a famous guitarist.

In this case, P1 is clearly false as a matter of fact, and the conclusion is also false. Nevertheless, the argument is valid: if both of the premises were true then the conclusion would also have to be true.

The above implies that a valid argument can have a false conclusion only if one or more of its premises is false. We say than an argument is **sound** just when it is valid *and* it has no false premises. By the definition of 'validity', if an argument is valid and all of its premises are true, then it must have a true conclusion. Therefore, since an argument is sound when it is valid and has no false premises, it must always have a true conclusion.

What about the following argument:

P1 Dogs exist.

C If Spike is a good dog, then Spike is a good dog.

Is this argument valid? Is it sound? We can see that it is valid by noting that the conclusion is always, necessarily, true. It therefore doesn't matter what the premises are: there is no possible way for all of the premises to be true and the conclusion false, because there's no way for the conclusion to be false *simpliciter*. The argument is also sound, since it is valid and it has no false premises.

> An argument is VALID just in case it is not possible for all the premises to be true and the conclusion false. An argument is SOUND just in case it is valid and it has no false premises.

We can think of *validity* and *soundness* as ideal standards for an argument: an argument that is valid and sound will have only true premises, and must have a true conclusion. Arguments don't get much better than that!

In Parts I and II of this text we will be focused on understanding validity. We focus on validity because doing so allows us to abstract away from particular matters of fact. Whether an argument is sound or unsound depends on the truth of its premises, so often we cannot determine whether an argument is sound without going out into the real world to check the truth of the premises. Validity, on the other hand, doesn't depend on whether the premises are *actually* true—what matters is just the logical relationship between the premises and the conclusion.

It is also important to keep in mind that there are many *in*valid and *un*sound arguments that nevertheless provide compelling reasons for accepting their conclusions. You will likely come across many such arguments in your study of philosophy, and many compelling arguments in the empirical sciences are probabilistic in nature rather than deductively valid. (Sometimes these are called 'inductively valid'.) Validity and soundness are very nice properties for an argument to have, but do not be too quick to dismiss an argument just because it isn't deductively valid!

3. Logical Form and Formal Validity

Compare the following two arguments:

P1 All humans are mortal.
P2 All Greeks are human.

C All Greeks are mortal.

P1 All hippos are morally inept.
P2 All galaxies are hippos.

C All galaxies are morally inept.

The argument on the right is unsound, because its second premise is false (and its first premise is questionable). Yet both arguments are valid, and they are valid for the same reason—they have the same **valid form**. Medieval logicians referred this particular form of argument as *Barbara*. (I'll explain why in a moment.) Such arguments go like this:

P1 All Hs are Ms.
P2 All Gs are Hs.

C All Gs are Ms.

In first argument, the letter 'H' stands for *human*, 'M' stands for *mortal*, and 'G' stands for *Greek*. In the second argument, 'H' stands for *hippos*, 'M' stands for *morally inept*, and 'G' stands for *galaxies*. Both arguments have the same form, and *every* argument that has this form must be valid. These arguments are **formally valid**; that is, you can tell that they are valid merely by considering their form, without even checking what the letters 'H', 'M', and 'G' stand for. **Formal logic** studies arguments with valid logical form. That is, it the study of which arguments have valid logical form and which arguments lack it, and what the difference between these two kinds of arguments amounts to.

Not every valid argument has a valid form. To see this, compare the following:

P1 Andrew is a bachelor.

C Andrew is a male.

P1 Andrew is a boat.

C Andrew is a magpie.

The argument on the left is valid: if Andrew is a bachelor then he must, by definition, be male, so if the premise is true then the conclusion must be true. The argument on the right, by contrast, is not valid. But the basic form of both arguments is the same:

P1 A is a B.

C A is a M.

The lesson here is that there can be valid arguments that are not formally valid—validity *does not* imply formal validity. On the other hand, every argument that has a valid form is therefore a valid argument—formal validity *does* imply validity. This means that if you can show that an argument has valid form, then you can know that it is valid without doing any further work. Establishing formal validity is a useful shortcut to establishing validity.

> An argument is FORMALLY VALID just in case, if any other argument shares its logical form, then it is a valid argument.

In order to characterise formal validity, we will need a way to characterise the **logical form** of an argument. This, in turn, requires a way of characterising the logical form of the sentences that make up the argument. We do this by translating our arguments into a **formal language**, which uses symbols to represent sentences and the parts thereof.

There are some formal languages that work like the symbolisation we gave for the two arguments we just looked at earlier in this section. A logic like this was developed by Aristotle in the 4th century BC. Aristotle's logic, with some revisions, was the dominant logic in the western world for more than two millennia. In the Aristotelian syllogistic logic, categories (like *human* and *mortal* and *vegetable*) are replaced with capital letters. Every line of an argument is then represented as having one of four forms, which medieval logicians labelled in this way:

A All As are Bs
E No As are Bs
I Some A is a B
O Some A is not a B

It is then possible to describe valid *syllogisms*, which are simple three-line arguments similar to *Barbara* above. Medieval logicians gave mnemonic names (like *Celarent*, *Baroco*, and *Felapton*) to all of the valid argument forms that could be created using sentences like these. As you can see, *Barbara* has the form **AAA**.

There are many limitations to Aristotelean logic. One is that it is very restricted in application. Aristotelian logic cannot represent a very large range of valid inferences that we can and often do make use of. For instance, the following argument is clearly valid (and sound), but is not an Aristotelean syllogism and cannot be represented in the Aristotelian syllogistic system:

P1 Either $1 + 1 = 2$, or $1 + 1 = 3$.
P2 If $1 + 1 = 3$, then $3 - 1 = 1$.
P3 It's not the case that $3 - 1 = 1$.

C $1 + 1 = 2$.

More generally, there are many sentences—and valid inferences involving them—that do not have a logical form that can be represented by the labels **A**, **E**, **I**, and **O**. This makes it impossible to apply the Aristotelian logic to arguments that make use of such sentences. Luckily for us, Aristotelean logic has been superseded, and modern logicians make use of formal languages with much greater expressive capacity.

4. Atomic and Non-Atomic Sentences

The remainder of Part I will develop a simple formal language, which we will label $\mathcal{L}_S$, and some tools for working out whether an argument formalised in this language is or is not formally valid. The subscript 'S' stands for 'sentential,' as $\mathcal{L}_S$ is the kind of language used for **sentential logic**. (Sentential logic is sometimes also called *propositional logic.*) In $\mathcal{L}_S$, we will use letters to represent atomic declarative sentences, and other special symbols to represent words like 'and' and 'or', which are used to link atomic sentences together to create more complicated sentences. This language will also be useful when we discuss probabilities in Part III.

In $\mathcal{L}_S$, italicised capital letters (e.g., A, B, C, . . ., Z) are always used to represent whole declarative sentences. We will refer to these as **sentence letters**. A sentence letter can in principle stand for any declarative sentence whatsoever. So, when translating from English into $\mathcal{L}_S$ (and *vice versa*), it is important to provide a **symbolisation key**. The key provides the intended reference for each sentence letter used in the symbolisation. For example, consider the following argument:

P1 Jack went for a swim.
P2 If Jack went for a swim, then Jack went to the river.

C Jack went to the river.

This is clearly a valid argument; you may recognise it as an instance of what's usually called *modus ponens*. What happens if we replace each premise in the argument with a single letter, according to the following symbolisation key:

A : Jack went for a swim.
B : If Jack went for a swim, then Jack went to the river.
C : Jack went to the river.

In that case, the argument will be:

P1 A
P2 B

C C

But… that's not very helpful! When we symbolise a valid argument, the goal should be to clearly represent the relevant structure of that argument in a way that makes it clear *why* the argument is valid, and we haven't yet done so here—there are many arguments that have the form 'A, B, therefore C' which aren't valid!

We can do better. Note first that the second premise of that argument is not just *any* sentence; rather, it is a non-atomic sentence that contains the first premise and the conclusion as parts. So our symbolisation key for the argument really only needs to include A and C,

A : Jack went for a swim.
C : Jack went to the river.

Given that, we can construct the second premise out of those two parts, like so:

P1 A
P2 If A, then C

C C

This is a *much* better representation of the structure of the argument, because it captures the relevant logical form of the argument in a way that is much more clear—you can immediately tell that any argument with this structure must be a valid argument. For instance, each of the following is clearly valid:

P1 Spot is a dog.
P2 If Spot is a dog, then Spot is friendly.

C Spot is friendly.

P1 Fluffy is a cat.
P2 If Fluffy is a cat, then Fluffy is evil.

C Fluffy is evil.

It will therefore be helpful to distinguish between simple (or **atomic**) sentences and complex (or **non-atomic**) sentences. An atomic sentence is a sentence that does not contain any other sentences as proper parts. Any non-atomic sentence contains at least one atomic sentence as a proper part. Atomic sentences are so-called because they are the basic building blocks of our language $\mathcal{L}_S$. For example, the sentence 'Roses are red' is atomic, whereas 'If roses are red then violets are blue' is not—the latter has two atomic sentences ('Roses are red' and 'Violets are blue') as parts, connected by 'If… then…'. The sentence 'It is not the case that roses are red' is also non-atomic: it contains 'Roses are red' as a proper part, with 'It is not the case that…' out the front.

An ATOMIC SENTENCE is a sentence that does not contain any other sentences as proper parts. Any non-atomic sentence contains at least one atomic sentence as a proper part.

The expression 'if… then…' is an example of a **connective**. Some other connectives include 'and', 'or', and 'not' (or 'it is not the case that…'). They are called 'connectives' because they connect (either atomic or non-atomic) sentences together to form new sentences. For example, suppose we have two atomic sentences, symbolised using the sentence letters 'A' and 'B'. We can then use these plus the connectives to construct more complex sentences, such as:

▷ A and B
▷ A and not B
▷ B or A
▷ If A, then B
▷ If A and not B, then not B

We will want to have symbols not only for sentences, but also for connectives. There are five different connectives that are symbolised in $\mathcal{L}_S$. The table that follows summarises them. Chapter 2 focuses on the first three connectives, $\neg$, $\wedge$ and $\vee$, while Chapter 3 focuses on the conditionals, $\rightarrow$ and $\leftrightarrow$.

symbol	what it is called	what it means
$\neg$	negation	'It is not the case that...'
$\wedge$	conjunction	'Both... and ...'
$\vee$	disjunction	'Either... or ...'
$\rightarrow$	conditional	'If ... then ...'
$\leftrightarrow$	biconditional	'... if and only if ...'

Finally, note that there are only twenty-six letters of the English alphabet, however we should want there to be no limits on the number of atomic sentences we might consider. To deal with this, we can use the same letter to symbolise different atomic sentences by adding subscripts. For example, we could have a symbolisation key that looks like this:

A_1 : The apple is under the armoire.
A_2 : Arguments in $\mathcal{L}_S$ always contain atomic sentences.
A_3 : Adam Ant is taking an airplane from Anchorage to Albany.
$\vdots$
A_{294} : Alliteration angers otherwise affable astronauts.

Subscripts may be numbers, as in the examples here, or other letters, or anything else whatsoever—what matters is just that the subscripts allow us to differentiate between sentence letters. Keep in mind that each of 'A_1', 'A_2', and so on, are to be considered a different sentence letter, even though they all use 'A'. When there are subscripts in the symbolisation key, it is therefore important to keep track of them.

Chapter 1: The Key Ideas

- ▷ An **argument** is an ordered sequence of two or more **propositions**, exactly one of which—the final one in the sequence—is the **conclusion**, and the rest of which are **premises**. An argument always has exactly one conclusion.

- ▷ A **proposition** is the meaning of, or what is expressed by, a **declarative sentence**. A declarative sentence is a sentence that is **truth-apt**. One proposition can be expressed by many different sentences, and in some special cases many propositions can be expressed by the same sentence in different contexts. Not all sentences express a proposition.

- ▷ There are exactly two **truth-values**, *true* and *false*. A sentence is true just in case the proposition it expresses is true, and a sentence is false just in case the proposition it expresses is false.

- ▷ An argument is **valid** when it is impossible for the premises to be true and the conclusion false. An argument is **sound** when it is valid and has no false premises. Every argument is either valid or invalid, and every argument is either sound or unsound.

- ▷ An argument is **formally valid** when its **logical form** guarantees its validity—that is, when any argument with the same logical form must be valid. In order to check for formal validity, we must be able to represent the logical form of an argument; this requires translating the argument into a **formal language**.

- ▷ The language of **sentential logic** is $\mathcal{L}_S$. In this language, we can use italicised capital English letters (A, B, C, . . .), sometimes with subscripts, to represent whole declarative sentences. These are called **sentence letters**. A **symbolisation key** tells us what sentence of English each sentence letter corresponds to.

- ▷ Usually, but not necessarily, a sentence letter will be used to represent an **atomic sentence**. An atomic sentence is a sentence that does not contain any other sentences as proper parts. We also use **connectives** ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$) to construct new sentences from one or more other sentences.

Practice Exercises

⋆ **Part A**
Which of the following sentences are declarative?

1. England is larger than China.
2. Greenland is south of Jerusalem.
3. Is New Jersey east of Wisconsin?
4. The atomic number of helium is 2.
5. The atomic number of helium is $3\pi^2$.
6. I hate seafood.
7. Blech—seafood, yuck!
8. Please hurry up.
9. Jack is over there.
10. Go over there.

⋆ **Part B**
Which of the following are true?

1. A valid argument can have one false premise and one true premise.
2. A valid argument can have a false conclusion.
3. A valid argument cannot have a false conclusion and all true premises.
4. Every valid argument is sound.
5. Every valid argument with true premises has a true conclusion.
6. Every unsound argument has a false conclusion.
7. Every argument with a false conclusion and only true premises is invalid.
8. There are sound arguments that are invalid.
9. There are arguments with a valid form that are invalid.
10. There are arguments with an invalid form that are valid.

Chapter 2

Negation, Conjunction, and Disjunction

This chapter introduces a fragment of the formal language $\mathcal{L}_S$. Three connectives, negation ($\neg$), conjunction ($\wedge$), and disjunction ($\vee$), will be introduced, and we will discuss how to translate ordinary English sentences into $\mathcal{L}_S$ using these connectives.

1. Negation

Consider how we might symbolise the following sentences of English:

1. Mary is in Barcelona.
2. Mary is not in Barcelona.
3. Mary is somewhere other than Barcelona.

Sentence 1 is an atomic sentence, so we can symbolise it using just one sentence letter. We can use this symbolisation key:

B : Mary is in Barcelona.

Sentence 2 is non-atomic. It is another way of saying 'It is not the case that Mary is in Barcelona', which contains Sentence 1 as a proper part. The 'It is not the case that…' part of the sentence is called a **negation**. In order to fully symbolise Sentence 2, we therefore need an appropriate symbol for negation. We will use $\neg$. Now we can symbolise Sentence 2 as $\neg B$. In this case, B is called the **negand**—it is the sentence being negated, which is always the sentence directly to the right of the negation symbol.

Sentence 3 does not contain the word 'not' anywhere. However, it *means* the same thing as Sentence 2. Both express the very same proposition. As such, because we symbolised Sentence 2 using $\neg B$, we can also symbolise Sentence 3 in the same way. Generally speaking, if two sentences that mean the same thing—if they express the same proposition—then they can be symbolised the same way. This gives us a general rule for using the negation symbol:

NEGATION: a sentence can be symbolised $\neg\mathcal{X}$ if it means the same thing as 'It is not the case that $\mathcal{X}$.'

In our statement of the rule for negation, we have used the fancy symbol $\mathcal{X}$. This is *not* a sentence letter. Rather, it is a *variable* that can be used to refer to any atomic *or* non-atomic sentence. So, $\mathcal{X}$ might refer to an atomic sentence A or B, or to a non-atomic sentences like $\neg B$, or to any of the other more complex symbolisations that we will be considering below. So, for example, the rule is telling us that if a sentence means the same thing as 'It is not the case that $\neg B$', then we can symbolise it as $\neg\neg B$.

Now consider these further examples:

4. The widget can be replaced.
5. The widget is irreplaceable.
6. The widget is not irreplaceable.

We can start with this symbolisation key:

R : The widget can be replaced.

In this case, Sentence 4 can be symbolised as just R. What about Sentence 5? Saying that the widget is irreplaceable means that it is not replaceable. So, Sentence 5 expresses the negation of Sentence 4, and since we've already symbolised Sentence 4 using R, we should symbolise Sentence 5 using the negation symbol as per our rule. This means that Sentence 5 can be symbolised as $\neg R$. This symbolisation lets us represent the important logical relationship between Sentence 4 and Sentence 5.

Sentence 6 is a little more tricky. It can be paraphrased as 'It is not the case that the widget is irreplaceable'. Since we have symbolised Sentence 5 as $\neg R$, in order to symbolise Sentence 6 we should use negation twice and represent it as $\neg\neg R$. This is a double-negation. In this case, $\neg R$ is the negand of the first negation symbol, and R is the negand of the second negation symbol.

That's enough for the symbolisation of negation. Let's now look at an interesting property that negation has. In particular, notice that for any sentence $\mathcal{X}$, if $\mathcal{X}$ is true then $\neg\mathcal{X}$ is false. And similarly, if $\mathcal{X}$ is false, then $\neg\mathcal{X}$ is true. Very simply, putting a $\neg$ out the front of any sentence $\mathcal{X}$ always gives us a new sentence with the opposite truth-value. This means that the truth-value of $\neg\mathcal{X}$, whatever it may be, is directly connected to the truth-value of $\mathcal{X}$: the former is a *function of* the latter.

Using the symbols T for *true* and F for *false*, we can summarize these truth-functional properties in what we'll call the **characteristic truth table** for negation. The truth table lays out in an easy to read manner exactly how the truth-value of $\neg\mathcal{X}$ is directly related to the truth-value of $\mathcal{X}$.

$\mathcal{X}$	$\neg\mathcal{X}$
T	F
F	T

The first row says that if $\mathcal{X}$ is true (T), then $\neg\mathcal{X}$ is false (F). The second row says that if $\mathcal{X}$ is false, then $\neg\mathcal{X}$ is true. You should find this intuitive, given what you already understand of how 'It is not the case that. . . ' works in English. (For reasons that we will get back to in Chapter 4, we will put the Ts and Fs underneath the $\neg$ symbol to represent the truth value of the whole sentence $\neg\mathcal{X}$.)

A brief point of clarification. Unlike the other connectives that we will discuss in this book (such as 'and' and 'or'), negation is not used to *connect* two different sentences together to form a bigger sentence. Instead, it simply creates a new sentence out of just *one* former sentence. Nevertheless, we will still refer to negation as a one-place sentential connective, because it shares interesting properties with the two-place connectives that we will discuss below—namely, it allows us to create non-atomic sentences out of atomic sentences, such that the truth-value of the newly created sentence is intimately connected to the truth-value(s) of the sentence(s) from which it is created.

2. Conjunction

Consider the following three sentences:

7. Adam is athletic.
8. Barbara is athletic.
9. Adam is athletic, and Barbara is also athletic.

We will need separate sentence letters to symbolise Sentence 7 and Sentence 8, so we use this symbolisation key:

A : Adam is athletic.
B : Barbara is athletic.

Sentence 7 can then be symbolised as simply A, and Sentence 8 can be symbolised as B. Given that, note that Sentence 9 can be paraphrased as simply 'A and B.' So, in order to symbolise Sentence 9, we need a symbol for the connective '... and...'. We will use $\wedge$, and we will symbolise 'A and B' as $A \wedge B$. The logical connective $\wedge$ is called **conjunction**, and the sentences it conjoins are called its **conjuncts**.

Notice that we have made no attempt to symbolise the word 'also' in Sentence 9. Words like 'both' and 'also' function in English to draw our attention to the fact that two things are being conjoined. However, they are not doing any further logical work: Sentence 9 means exactly the same thing as 'Adam is athletic, and Barbara is athletic.' So, we do not need to represent words like 'both' and 'also' in our symbolic language $\mathcal{L}_S$. We are trying to capture the *relevant logical structure* of the sentences, not their precise structure in English.

CONJUNCTION: a sentence can be symbolised $\mathcal{X} \wedge \mathcal{Y}$ if it means the same thing as '$\mathcal{X}$ and $\mathcal{Y}$.'

Some more examples:

10. Barbara is athletic and energetic.
11. Barbara and Adam are both athletic.
12. Although Barbara is energetic, she is not athletic.
13. Barbara is athletic, but Adam is more athletic than she is.

Sentence 10 is a conjunction. It says two things about Barbara, so in English it is permissible to refer to Barbara only once. It might be tempting to try this when symbolising Sentence 10: since B means 'Barbara is athletic', one might paraphrase the sentence as 'B and energetic.' But this would be a mistake: the conjuncts of a conjunction must always be whole sentences, and 'energetic' by itself is not a whole sentence. We should instead introduce a new sentence letter C to stand for 'Barbara is energetic', and symbolise Sentence 10 as $B \wedge C$.

Sentence 11 says one thing about two different subjects. It says of both Barbara and Adam that they are each athletic, and in English when we say this we may use the word 'athletic' only once as with this sentence. However, when symbolising in $\mathcal{L}_S$, it is important to realise that the Sentence 11 means the same thing as 'Barbara is athletic, and Adam is athletic.' Hence we should symbolise Sentence 11 as $B \wedge A$.

Sentence 12 is a bit more complicated. The word 'although' sets up a contrast between the first part of the sentence and the second part. Nevertheless, it says both that Barbara is energetic and that she is not athletic. In order to make each of the conjuncts an atomic sentence, we need to replace 'she' with 'Barbara.' So we can paraphrase Sentence 12 as 'Barbara is energetic, and Barbara is not athletic.' The second conjunct contains a negation, so we paraphrase further: 'Barbara is energetic, and it is not the case that Barbara is athletic.' So the overall symbolisation of Sentence 12 is $C \wedge \neg B$.

Sentence 13 contains a similar contrastive structure: the word 'but' is usually used to indicate a contrast between two clauses in a complex sentence. Nevertheless, this contrast is irrelevant for the purposes of translating into $\mathcal{L}_S$, so we can paraphrase Sentence 13 as 'Both Barbara is athletic, and Adam is more athletic than Barbara.' (Notice that we once again replace the pronoun 'she' with 'Barbara.') How should we symbolise the second conjunct? We already have the sentence letter A which is about Adam's being athletic, and B which is about Barbara's being athletic, but neither is about one of them being *more* athletic than the other. So we need a new sentence letter. Let D be our symbol for the atomic sentence 'Adam is more athletic than Barbara.' Now Sentence 13 gets symbolised as $B \wedge D$. So here is an additional, supplementary rule for conjunction:

> BUT, ALTHOUGH: sentences that can be paraphrased as '$\mathcal{X}$, but $\mathcal{Y}$', or as 'Although $\mathcal{X}$, $\mathcal{Y}$', can also be symbolised as $\mathcal{X} \wedge \mathcal{Y}$.

One last example:

14. Barbara and Adam lifted the jukebox together.

How should we symbolise Sentence 14? You might be tempted to think that it is a conjunction, because of the presence of the word 'and.' However, note that the sentence *doesn't* seem to mean the same thing as 'Barbara lifted the jukebox and Adam lifted the jukebox', as this would usually be taken to imply that Barbara and Adam each lifted the jukebox *separately* (i.e., at different times). Sentence 14 is actually an atomic sentence, rather than a conjunction. It's important to keep this in mind: a declarative sentence only expresses a conjunction *if* it can be paraphrased as 'Both $\mathcal{X}$ and $\mathcal{Y}$.' So, sometimes a sentence can use the word 'and' but not be a conjunction.

Like negation, conjunction also has special truth-functional properties. To see this, note that for any two sentences $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{X} \wedge \mathcal{Y}$ is true just in case both $\mathcal{X}$ is true and $\mathcal{Y}$ is true. This means that $\wedge$ is a **truth-functional connective**: the truth-value of $\mathcal{X} \wedge \mathcal{Y}$ is a function of the truth-values of its two parts, $\mathcal{X}$ and $\mathcal{Y}$. Consequently, if you know the truth-values of the parts, you can work out the truth-value of any conjunction containing those parts. Here is the characteristic truth table for conjunction:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \wedge \mathcal{Y}$
T	T	T
T	F	F
F	T	F
F	F	F

In this case, the first row says that if $\mathcal{X}$ is true and $\mathcal{Y}$ is true, then $\mathcal{X} \wedge \mathcal{Y}$ is true. The second row says that if $\mathcal{X}$ is true and $\mathcal{Y}$ is false, then'$\mathcal{X} \wedge \mathcal{Y}$ is false. And so on. The truth table as a whole tells us that there's only one way for a conjunction to be true: if both of its conjuncts are true. In all other cases the conjunction is false.

Note that conjunction is **symmetrical**, in the sense that we can swap the order of the conjuncts without any changes in the truth-value of the conjunction overall. That is, regardless of what $\mathcal{X}$ and $\mathcal{Y}$ are, $\mathcal{X} \wedge \mathcal{Y}$ is true just in case $\mathcal{Y} \wedge \mathcal{X}$ is true. Only the truth-values of the conjuncts makes a difference, not the order of those conjuncts in the conjunction. This is an important property of conjunction and several of the other connectives that we will consider—but not all of them.

3. Disjunction

Two new sentences to consider:

15. Either Denison will play golf, or he will watch movies.
16. Either Denison or Ellery will play golf.

We will use this symbolisation key:

D : Denison will play golf.
E : Ellery will play golf.
M : Denison will watch movies.

Sentence 15 has the basic form 'Either D or M.' To fully symbolise this we will want a new symbol for 'or'. For this we'll use $\vee$. Thus Sentence 15 becomes $D \vee M$. The connective '$\vee$' is called **disjunction**, and the sentences it connects are its **disjuncts**.

Sentence 16 is only slightly more complicated. There are two subjects, but the English sentence only gives the verb once. In symbolizing, we can paraphrase it as 'Either Denison will play golf, or Ellery will play golf.' Hence, it can be symbolised as $D \vee E$.

DISJUNCTION: a sentence can be symbolised $\mathcal{X} \vee \mathcal{Y}$ if it means the same thing as 'Either $\mathcal{X}$, or $\mathcal{Y}$.'

Sometimes in English, the word 'or' can be plausibly understood as excluding the possibility that both disjuncts are true. This is sometimes called an **exclusive or**. For instance, an exclusive or is clearly intended when it says, on a restaurant menu, 'Entrees come with either soup or salad.' You may have soup, or you may have salad—but if you want *both*, then you have to pay extra. At other times, though, the use of 'or' allows for the possibility that both disjuncts might be true. This is probably the case with Sentence 16, above. Denison might play, or Ellery, or both might. Sentence 16 merely says that *at least* one of them will play golf. This is called an **inclusive or**. (If it helps, you might think of the inclusive or as always having an implicit 'or both' attached to it.) The symbol $\vee$ in our formal language will always represents an inclusive or. This is very important to keep in mind—many a logic student has run into error as a result of confusing the inclusive and exclusive or. So $\mathcal{X} \vee \mathcal{Y}$ is true if $\mathcal{X}$ is true, if $\mathcal{Y}$ is true, or if *both* $\mathcal{X}$ and $\mathcal{Y}$ are true. This means that it is false *only* if both $\mathcal{X}$ and $\mathcal{Y}$ are false.

Disjunction is, therefore, another kind of truth-function: for any two sentences $\mathcal{X}$ and $\mathcal{Y}$, if you know the truth-values of $\mathcal{X}$ and $\mathcal{Y}$ then you can work out the truth value of $\mathcal{X} \vee \mathcal{Y}$. Like conjunction, disjunction is symmetrical: $\mathcal{X} \vee \mathcal{Y}$ is just another way of saying $\mathcal{Y} \vee \mathcal{X}$, and they always have the same truth-value. We can summarise this with the characteristic truth table for disjunction:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \vee \mathcal{Y}$
T	T	T
T	F	T
F	T	T
F	F	F

4. Putting it all together

We have considered translating $\neg$, & , and $\vee$ sentences separately; now it's time to consider translating sentences that use these in different combinations. So consider:

17. Either you will not have soup, or you will not have salad.
18. You will have neither soup nor salad.
19. You will have soup or salad, and a sweet surprise.
20. You get either soup or salad, but not both.

We will use the following key:

S_1 : You will have soup.
S_2 : You will have salad.
S_3 : You will have a sweet surprise.

Then, Sentence 17 can be paraphrased in this way: 'Either it is not the case that you will have soup, or it is not the case that you will have salad.' Translating this requires both disjunction and negation. It becomes $\neg S_1 \vee \neg S_2$.

Sentence 18 also requires negation. It can be paraphrased as, 'It is not the case that: either you will have soup or you will have salad.' To symbolise this we will need some way of indicating that the negation does not only negate the right or the left disjunct, but instead negates the entire disjunction. That is, the **scope** of the negation symbol $\neg$ in this case is the entire disjunction. In order to faithfully represent the logical form of Sentence 18, we can put parentheses around the disjunction: 'It is not the case that: $(S_1 \vee S_2)$.' So the proper symbolisation of Sentence 18 becomes simply $\neg(S_1 \vee S_2)$.

Notice that the parentheses are doing very important work here. If we tried to symbolise Sentence 18 as just $\neg S_1 \vee S_2$, without the parentheses, then it could be taken to mean 'Either it is not the case that you will have soup, or you will have salad.' In this case we would say that the scope of the negation is limited to just S_1, and this is not what we want. $\neg S_1 \vee S_2$ and $\neg(S_1 \vee S_2)$ are very different sentences with very different meanings. The same is true, for example, of $\neg S_1 \wedge S_2$ and $\neg(S_1 \wedge S_2)$. In general, the correct placement of parentheses is critically important for converting complex sentences into symbolic form.

There is another way you might symbolise Sentence 18. Above, we paraphrased it as 'It is not the case that: either you will have soup or you will have salad.' But we could also have paraphrased it as 'It is not the case that you get soup, and it is not the case that you will have salad.' In that case, we could instead have symbolised Sentence 18 as $\neg S_1 \wedge \neg S_2$. This is not a problem—you can legitimately symbolise Sentence 18 as either $\neg(S_1 \vee S_2)$ or $\neg S_1 \wedge \neg S_2$, as they are really just different ways of saying the same thing. (You will be able to check this for yourself after you learn about truth tables in the next chapter.)

The parentheses are also very important for symbolising Sentence 19. Suppose we just wrote it down as $S_1 \vee S_2 \wedge S_3$. This would be ambiguous between:

(i) You will have soup or salad, and you will have a sweet surprise.
(ii) Either you will have soup, or you will have salad and a sweet surprise.

If we use parentheses, however, we can disambiguate between these. The former disambiguation becomes $(S_1 \vee S_2) \wedge S_3$, while the latter becomes $S_1 \vee (S_2 \wedge S_3)$. Again, these are very different sentences with very different meanings, so it is very important to make sure you know how parentheses are properly used in logical symbolisation.

Finally, despite its relatively simple expression in English, Sentence 20 is surprisingly complex when symbolised in $\mathcal{L}_S$. We can break the sentence down into two parts. The first part says that you get one or the other of soup or salad. This is a simple disjunction which we can symbolise as $S_1 \vee S_2$. Recall that this is an inclusive or, so the first part of Sentence 20 is consistent with 'or both'. The second part of Sentence 20 then rules this out, by explicitly saying that you cannot have both. We can paraphrase the second part as 'It is not the case that you get soup and you get salad.' Note that the negation is here taking scope over the entire conjunction, so it should be symbolised as $\neg(S_1 \wedge S_2)$ rather than $\neg S_1 \wedge S_2$. (The latter would say that you don't get soup, but you do get salad.)

Now we just need to put the two parts, $S_1 \vee S_2$ and $\neg(S_1 \wedge S_2)$, together. As we saw above, 'but' can usually be symbolised as a conjunction, so we'll use $\wedge$ to join the two parts. Since $S_1 \vee S_2$ is one whole conjunct in the conjunction, we'll need to put parentheses around it to avoid ambiguity. So, finally, Sentence 20 can thus be symbolised as

$$(S_1 \vee S_2) \wedge \neg(S_1 \wedge S_2)$$

In other words: you can have soup or you can have salad, and you cannot have both soup and salad. Despite its relatively simple appearance in English, Sentence 20 is in $\mathcal{L}_S$ represented by something rather more complicated—a conjunction of a disjunction and the negation of a conjunction.

There is an important philosophical lesson here: sometimes, simple sentences in English might express logically complex meanings. It can take quite some insight to determine the right way to translate an English sentence into logical notation.

Chapter 2: The Key Ideas

- The symbol $\neg$ represents **negation**. In a sentence of the form $\neg\mathcal{X}$, $\mathcal{X}$ is called the **negand**. A sentence can be symbolised as $\neg\mathcal{X}$ if it means the same thing as 'It is not the case that $\mathcal{X}$.'

- The symbol $\wedge$ represents **conjunction**. In the sentence of the form $\mathcal{X} \wedge \mathcal{Y}$, the sentences $\mathcal{X}$ and $\mathcal{Y}$ are called the **conjuncts**. A sentence can be symbolised as $\mathcal{X} \wedge \mathcal{Y}$ if it means the same thing as 'Both $\mathcal{X}$, and $\mathcal{Y}$.' Sentences that can be paraphrased as '$\mathcal{X}$, but $\mathcal{Y}$' or 'Although $\mathcal{X}$, $\mathcal{Y}$' can also be symbolised as $\mathcal{X} \wedge \mathcal{Y}$. Be careful: not all sentences that have 'and' in them are conjunctions.

- The symbol $\vee$ represents **disjunction**. In the sentence of the form $\mathcal{X} \vee \mathcal{Y}$, the sentences $\mathcal{X}$ and $\mathcal{Y}$ are called **disjuncts**. A sentence can be symbolised as $\mathcal{X} \vee \mathcal{Y}$ if it means the same thing as 'Either $\mathcal{X}$, or $\mathcal{Y}$.' The language $\mathcal{L}_S$ uses an **inclusive or**. In other words, a sentence of the form $\mathcal{X} \vee \mathcal{Y}$ is always considered true when either of the disjuncts are true, including when both are true.

- $\neg$, $\wedge$ and $\vee$ are all **truth-functional connectives**. This means that the truth-values of the sentence created using those connectives will always be a function of the truth-values of its parts. For example, the truth-value of the negation $\neg A$ depends on the truth-value of its negand, A. Similarly, the truth-value of the conjunction $A \wedge B$ depends on the truth-value of its conjuncts, A and B. This dependence is represented by a **characteristic truth table**:

$\mathcal{X}$	$\neg\mathcal{X}$
T	F
F	T

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \wedge \mathcal{Y}$
T	T	T
T	F	F
F	T	F
F	F	F

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \vee \mathcal{Y}$
T	T	T
T	F	T
F	T	T
F	F	F

Practice Exercises

⋆ **Part A**
Using the symbolisation key provided, symbolise each of the following into its nearest $\mathcal{L}_S$ equivalent.

M : Those creatures are men in suits.
C : Those creatures are chimpanzees.
G : Those creatures are gorillas.

1. Those creatures are not men in suits.
2. Those creatures are gorillas, not chimpanzees.
3. Those creatures are men in suits, or they are not.
4. Those creatures are either gorillas or chimpanzees.
5. Those creatures are neither gorillas nor chimpanzees.
6. Although those creatures are not men in suits, they are not gorillas either.
7. Those creatures are gorillas or chimpanzees, but not both.

Part B
What is the *logical* difference between the sentences in the following pairs? How would you symbolise each one? (Choose your own symbolisation key.)

a. Alice and Charlie are fighters.
b. Alice and Charlie are fighting.

a. Charlie is not happy.
b. Charlie is unhappy.

Part C
Using the symbolisation key provided, symbolise each of the following into its nearest $\mathcal{L}_S$ equivalent.

E_a : Ava is an electrician.
E_h : Harrison is an electrician.
F_a : Ava is a firefighter.
F_h : Harrison is a firefighter.
S_a : Ava is satisfied with her career.
S_h : Harrison is satisfied with his career.

1. Ava and Harrison are both electricians.
2. Harrison is an unsatisfied electrician.
3. Neither Ava nor Harrison is an electrician.
4. Both Ava and Harrison are electricians, but neither find it satisfying.
5. Ava is satisfied with her career but Harrison is not satisfied with his.
6. It cannot be that Harrison is both an electrician and a firefighter.
7. Harrison and Ava are both firefighters, and neither is an electrician.
8. Although Ava is a firefighter, she is not satisfied with her career.

⋆ Part D

Provide a symbolisation key and then symbolise each of the following into its nearest $\mathcal{L}_S$ equivalent.

1. Alice and Bob each lifted the couch.
2. Either Alice or Bob lifted the couch, but not both.
3. Alice and Bob together lifted the couch.
4. Although Alice didn't lift the couch, Bob did.
5. Neither Alice nor Bob lifted the couch.

⋆ Part E

Which of the following pairs of sentences say the same thing, and which do not? Provide appropriate symbolisations for each.

1. Alice and Charlie are both playing football.
2. Both Alice and Charlie are playing football.

1. Alice and Charlie are not both playing football.
2. Alice and Charlie are both not playing football.

1. Either Alice or Charlie is playing football.
2. Neither Alice nor Charlie is playing football.

1. Neither Alice nor Charlie is playing football.
2. Alice is not playing football, and Charlie is not playing football either.

1. Either Alice is not playing football, or Charlie isn't.
2. It's not the case that Alice or Charlie is playing football.

Part F

Consider the following sentence: 'One mistake, John, and you will be fired.' Suppose:

M : John makes a mistake.
F : John will be fired.

Why would it be *incorrect* to symbolise the sentence as $M \wedge F$? Do you think there is an appropriate way to symbolise the sentence using just the symbols introduced in this chapter ($\neg$, $\wedge$, $\vee$)?

Chapter 3

Conditionals

This chapter adds two more truth-functional connectives to our simple formal language $\mathcal{L}_S$. These are the material conditional and the biconditional. At the end of the chapter, we will also give a complete characterisation of what it is to for something to count as a sentence in the language of $\mathcal{L}_S$.

1. Material Conditionals

Start with this symbolisation key:

R : You cut the red wire.
B : The bomb will explode.

Now consider the following sentence:

1. If you cut the red wire, then the bomb will explode.
2. The bomb will explode only if you cut the red wire.

Sentence 1 is a **conditional**, and it can be partially symbolised as 'If R, then B.' We can symbolise the 'If. . . , then. . . ' using an arrow $\rightarrow$. Given that, we can then symbolise Sentence 1 as $R \rightarrow B$. As you can see, a conditional sentence has two sub-sentences linked by an arrow. The sentence on the left-hand side of the arrow is called the **antecedent**. The sentence on the right-hand side of the arrow is called the **consequent**.

(Try not to confuse *antecedents* and *premises*. Likewise, try not to confuse *consequents* and *conclusions*. Antecedents and consequents are parts of a conditional; premises and conclusions are parts of arguments—they are not the same things, and mixing them up can sometimes cause a great deal of needless confusion.)

Sentence 2 is also a conditional. But be careful! Because the word 'if' appears in the second half of the Sentence 2, it might be tempting to symbolise this the same way that we symbolised Sentence 1. However, doing so would be a mistake. Sentence 1 says 'If R, then B.' In other words, R being true is a **sufficient condition** for the truth of B—the truth of R is sufficient to guarantee the truth of B. Notice, though, that Sentence 1 does *not* imply that your cutting the red wire is the *only* way that the bomb explodes. For example, even if you don't cut the wire, someone else might cut it, or the bomb might be on a timer and explode anyway regardless of whether the wire is cut. So Sentence 1 does not say anything about what will be the case if R is *false*.

On the other hand, Sentence 2 says that the *only* circumstance under which the bomb will explode involves your having cut the red wire. So Sentence 2 says that R is a **necessary condition** for the truth of B. Another way to say the same thing is that if B is true, then R must be true as well. Given this, Sentence 2 should really be symbolised as $B \to R$. In this case, we say that Sentence 2 is the **converse** of Sentence 1, and vice versa. That just means that their antecedents and consequents have been swapped.

Summarising the above, a sentence of the form 'If $\mathcal{X}$, then $\mathcal{Y}$' says that the truth of $\mathcal{X}$ is sufficient for the truth of $\mathcal{Y}$, or (equivalently) that the truth of $\mathcal{Y}$ is necessary for the truth of $\mathcal{X}$. (They are just two ways of saying the same thing.) So when you're trying to translate what appears to be a conditional sentence into our logical language, it helps to think: *Which half of the sentence is the sufficient condition for the other, and which half is the necessary condition for the other?* The sufficient condition should always go on the left of the arrow, and the necessary condition on the right.

Notice also that from what we've said about the correct way to symbolise Sentence 2, it follows that the following Sentence 3 says the same thing as Sentence 1:

3. You cut the red wire only if the bomb will explode.

Both Sentence 1 and Sentence 3 state that cutting the red wire is sufficient for the bomb to explode, or equivalently, that the bomb's exploding is a necessary condition of the red wire being cut. In other words, a sentence of the form 'If R, then B' means the same as a sentence of the form 'R only if B'. These are just different ways of writing the same thing. Hence our first rule for translating conditional sentences:

> MATERIAL CONDITIONAL: a sentence can be symbolised $\mathcal{X} \to \mathcal{Y}$ if it means the same thing as 'If $\mathcal{X}$, then $\mathcal{Y}$', or if it means the same thing as '$\mathcal{X}$ only if $\mathcal{Y}$'.

$\mathcal{X} \to \mathcal{Y}$ means that if $\mathcal{X}$ is true then so is $\mathcal{Y}$. It follows naturally, then, that if the antecedent $\mathcal{X}$ is true and the consequent $\mathcal{Y}$ is false, then the whole conditional must be false. So one way for the conditional to be false is for its antecedent to be true and its consequent to be false. We can represent this in the following partial truth table:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \to \mathcal{Y}$
T	T	
T	F	F
F	T	
F	F	

But what is the truth value of '$\mathcal{X} \to \mathcal{Y}$' under the other circumstances? That's where conditionals can really start to get tricky. To begin, consider the sentence 'If the moon were made of cheese, then it would be cheddar.' This has an 'If . . . , then. . . ' structure, so it is kind of conditional, but it is a special kind of conditional. In philosophy we usually say that it is a **counterfactual conditional**, since it talks about what *would* be the case if something that's not in fact true *were* to be true (i.e., if things were counter-to-the-facts). To determine the truth of these sorts of conditionals, we usually have to try to work out how things would be different if the antecedent were to somehow be true.

There is a great deal of academic discussion about how exactly that process is supposed to work. But *we* don't have to worry about that, since in this text we are not going to be dealing with counterfactual conditionals. Instead, our symbol $\rightarrow$ represents the **material conditional**. By stipulation, a material conditional is always true whenever its antecedent is false. Logicians in this case will often say that the conditional is **trivially true**. Thus we can add two more rows to the foregoing truth table:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \rightarrow \mathcal{Y}$
T	T	
T	F	F
F	T	T
F	F	T

For example, the sentence 'If $1 + 1 = 3$, then the moon is made of cheese' is trivially true, when translated as a material conditional. And the sentence 'If $1 + 1 = 3$, then the moon is not made of cheese' is *also* trivially true, for the same reason. This is counterintuitive, to be sure, and sentences like these are instances of what's often called the *paradoxes of material implication*. Nevertheless, that is how $\rightarrow$ is defined.

This also means that it would *not* really be appropriate to symbolise the sentence 'If the moon were made of cheese, then it would be cheddar' using $\rightarrow$. Since that is a counterfactual conditional, it is *not* automatically true whenever the antecedent is false. In other words, the counterfactual conditional is not **truth-functional**, since the truth of the counterfactual conditional as a whole does not depend just on the truth or falsity of the antecedent and consequent—it also depends on how the antecedent and consequent are connected to one another. In fact, we cannot quite capture the meaning of counterfactual conditionals in our simple logical language, which only has truth-functional connectives. This isn't necessarily a bad thing, as it just means that our language has some limitations when it comes to translating from sentences of English. If you choose to go on to do more advanced studies in logic, you will eventually come across more complex logical languages that are better equipped to handle counterfactuals.

Next, consider the sentence 'If the moon is round, then London is a city'. In this case, the antecedent is true and the consequent is also true, but there is no apparent connection between the two—it isn't the case that the roundness of the moon *makes* London a city. Nevertheless, if the sentence is translated using the material conditional $\rightarrow$, then we say that it is true whenever both the antecedent and consequent are true. This gives us the complete truth table:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \rightarrow \mathcal{Y}$
T	T	T
T	F	F
F	T	T
F	F	T

In short: there is exactly one circumstance in which a sentence of the form $\mathcal{X} \rightarrow \mathcal{Y}$ is said to be false, which is whenever $\mathcal{X}$ is true and $\mathcal{Y}$ is false. In all other circumstances $\mathcal{X} \rightarrow \mathcal{Y}$ is true. Notice that, unlike disjunction and conjunction earlier, the material conditional is **asymmetrical**. You cannot in general swap the antecedent and consequent without changing the meaning of the sentence, because $A \rightarrow B$ and $B \rightarrow A$ do not mean the same thing. It is important to make sure you get the direction of the arrow correct.

'Unless' sentences

Consider these sentences that use the word 'unless':

4. Unless you wear a jacket, you will catch cold.
5. You will catch cold unless you wear a jacket.

Use the following symbolisation key:

J : You will wear a jacket.
D : You will catch a cold.

We can now partially formalise Sentence 4 as 'Unless J, D.' This means that if you do not wear a jacket, then you will catch cold. With this in mind, we might translate it as $\neg J \to D$. It also means that if you do not catch a cold, then you must have worn a jacket; with this in mind, we might translate it as $\neg D \to J$. Which of these is the correct translation of Sentence 4? The answer is both—the two translations are logically equivalent in $\mathcal{L}_S$. (You will be able to prove this for yourself after reading the next chapter.) Sentence 5 is logically equivalent to Sentence 4. So it too can be translated either as $\neg J \to D$, or as $\neg D \to J$.

When symbolising sentences like Sentence 4 and Sentence 5, it is very easy to get turned around. Since the conditional is not symmetric, it would be wrong to translate either as $J \to \neg D$. Fortunately, there are other logically-equivalent expressions we could use as well. Both Sentence 4 and Sentence 5 mean that you will wear a jacket or—if you do not wear a jacket—then you will catch a cold. So we can also just translate them as $J \vee D$. (You might worry that the 'or' here should be an *exclusive or*. However, Sentence 4 and Sentence 5 do not exclude the possibility that you might *both* wear a jacket *and* catch a cold; jackets do not protect you from all the possible ways that you might catch a cold. So the use of an *inclusive or*, represented by $\vee$, is appropriate.)

Thus the general rule for translating 'unless' sentences:

UNLESS: a sentence of the form 'Unless $\mathcal{X}$, $\mathcal{Y}$' can be symbolised either as $\neg\mathcal{X} \to \mathcal{Y}$, or as $\neg\mathcal{Y} \to \mathcal{X}$, or as $\mathcal{X} \vee \mathcal{Y}$ (they all mean the same thing).

2. Biconditionals

Consider these sentences:

6. That figure is a triangle only if it has exactly three sides.
7. That figure is a triangle if it has exactly three sides.
8. That figure is a triangle if and only if it has exactly three sides.

We'll use this symbolisation key for the atomic sentences:

T : That figure is a triangle.
S : That figure has exactly three sides.

Sentence 6 can be symbolised as $T \to S$. On the other hand, Sentence 7 is importantly different. It can be paraphrased as, 'If the figure has three sides, then it is a triangle.' So it should be symbolised as $S \to T$. Sentence 8 says that T is true *if and only if* S is true. So we can infer S from T, and we can infer T from S. This is called a **biconditional**, because it entails the two conditionals $S \to T$ and $T \to S$. We will use $\leftrightarrow$ to represent the biconditional, so Sentence 8 can be symbolised as $S \leftrightarrow T$.

As we said in the last section, whenever $S \to T$ is true, then the truth of S is a sufficient condition for the truth of T, and the truth of T is a necessary condition for the truth of S. That means that, whenever $S \leftrightarrow T$ is true, the truth of S must be *both* a sufficient *and* necessary condition for the truth of T. Likewise, the truth of T must be necessary and sufficient for the truth of S. In other words, $S \leftrightarrow T$ says that S and T are necessary and sufficient *for each other*. You can't have either one being true without the other one being true, and if one of them is false then they both have to be false.

Because of this, $\leftrightarrow$ is also a truth-functional connective. Simply, any sentence of the form $\mathcal{X} \leftrightarrow \mathcal{Y}$ is true just in case either both of $\mathcal{X}$ and $\mathcal{Y}$ are true, or both are false. Another way to put that is to say that $\mathcal{X} \leftrightarrow \mathcal{Y}$ is true just in case $\mathcal{X}$ and $\mathcal{Y}$ *have the same truth-value.* Hence, here is the characteristic truth table for the biconditional:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \leftrightarrow \mathcal{Y}$
T	T	T
T	F	F
F	T	F
F	F	T

Of course, if we wanted to, we could abide without any new symbols for the biconditional. Since Sentence 8 just means 'If that figure is a triangle then that figure has exactly three side, *and* if that figure has exactly three sides, then that figure is a triangle', we could symbolise it as $(T \to S) \wedge (S \to T)$. So we do not strictly speaking *need* to introduce a new symbol for the biconditional. Nevertheless, logical languages usually have symbols for material conditionals and biconditionals for the sake of simplicity—and as we'll see in the next chapter, the biconditional is very important and useful in many cases.

3. Sentences of $\mathcal{L}_S$

The sentence 'It's not true that either apples are blue, or berries are pink' is a sentence of English, whereas $\neg(A \vee B)$ is a sentence of our formal language $\mathcal{L}_S$. Although we can identify sentences of English when we encounter them, we do not yet have a precise and formal definition of 'sentence of English'. We can, however, have a precise and formal definition of what counts as a sentence of $\mathcal{L}_S$. This is one respect in which a formal language like $\mathcal{L}_S$ is more precise than a natural language like English.

We have seen that there are three basic kinds of symbols in $\mathcal{L}_S$:

Atomic sentences	$A, B, C, \ldots, Z$
with subscripts, as needed	$A_1, B_1, Z_1, A_2, A_{25}, J_{375}, \ldots$
Connectives	$\neg, \wedge, \vee, \to, \leftrightarrow$
Brackets	(,)

We can define an **expression of** $\mathcal{L}_S$ as *any* string of symbols of $\mathcal{L}_S$. That is, if you take any of the above symbols and write them down in any order, then you will have an expression of $\mathcal{L}_S$. Of course, most expressions of $\mathcal{L}_S$ in this sense will be total gibberish: for example, $() \wedge A)(\rightarrow)B \rightarrow\rightarrow \neg \rightarrow (AB))$ would not symbolise anything meaningful. This means that an expression of $\mathcal{L}_S$ is very different from a **sentence of** $\mathcal{L}_S$. We want to know when an expression of $\mathcal{L}_S$ amounts to a sentence of $\mathcal{L}_S$.

Clearly, our symbols for individual atomic sentences should all count as complete sentences of $\mathcal{L}_S$. So every sentence letter by itself is a sentence of $\mathcal{L}_S$. We can then create further sentences out of these by using the various connectives we've defined: $\neg$, $\vee$, &, $\rightarrow$, $\leftrightarrow$. Using negation, we can construct the new sentences $\neg A$ and $\neg B$ using the sentence letters A and B as a starting point. Using conjunction, we can get $(A \wedge B)$, $(B \wedge A)$, $(A \wedge A)$, and $(B \wedge B)$, and so on. We could also apply negation repeatedly to get sentences like $\neg\neg A$, $\neg\neg\neg A$, $\neg\neg\neg\neg A$. Or we could apply negation along with conjunction to get sentences like '$\neg(A \wedge B)$ and $\neg(B \wedge \neg B)$. All of these are sentences of $\mathcal{L}_S$.

As you can see, the possible combinations are endless—even starting with just the two sentence letters A and B, there are infinitely many ways to create new sentences from them using the connectives we've considered. And there are infinitely many sentence letters, so even more combinations still. Because of this, there's no point in trying to list all the possible sentences one by one. Instead, we will describe the process by which sentences can be systematically *constructed* according to a system of rules for constructing new sentences out of old sentences plus connectives.

Consider such a rule for negation. It will say that for any sentence $\mathcal{X}$ of $\mathcal{L}_S$, $\neg\mathcal{X}$ will also be a sentence of $\mathcal{L}_S$. So if A is a sentence, then $\neg A$ is a sentence too. And if $\neg(A \wedge B)$ is a sentence, then $\neg\neg(A \wedge B)$ is a sentence too. That is a very simple rule that we can apply to anything we already know is a sentence, in order to create a new sentence. We can say similar things for each of the other connectives. For instance, if $\mathcal{X}$ and $\mathcal{Y}$ are both sentences of $\mathcal{L}_S$, then $(\mathcal{X} \wedge \mathcal{Y})$ is also a sentence of $\mathcal{L}_S$. Providing clauses like this for all of the connectives, we arrive at the following complete and precise definition of what it is for an expression to count as a sentence of $\mathcal{L}_S$:

Definition of a SENTENCE OF $\mathcal{L}_S$:

1. Every atomic sentence is a sentence.
2. If $\mathcal{X}$ is a sentence, then $\neg\mathcal{X}$ is a sentence.
3. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \wedge \mathcal{Y})$ is a sentence.
4. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \vee \mathcal{Y})$ is a sentence.
5. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \rightarrow \mathcal{Y})$ is a sentence.
6. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \leftrightarrow \mathcal{Y})$ is a sentence.
7. Nothing else is a sentence.

Definitions like this are called **recursive**. Recursive definitions begin with some specifiable *base elements*—in this case the set of atomic sentences—and then give us ways to generate indefinitely many more *derivative elements* by compounding together previously established elements according to certain rules. To give you a better idea of what a recursive definition is, we can give a recursive definition of the idea of an *ancestor*. We specify a base clause:

- x's parents are x's ancestors

and then offer further clauses like:

- If y is an ancestor of x, then y's parents are also x's ancestors
- Nothing else is an ancestor of x

Using this definition, we can easily check to see whether someone is Sally's ancestor. If the person is one of Sally's parents, or a parent of Sally's parents, or..., then the person is Sally's ancestor; otherwise, they are not one of Sally's ancestors. By specifying the base elements (Sally's parents are her ancestors), plus a rule (if someone is a parent of an ancestor of Sally, then they are an ancestor of Sally too), we have in effect compressed an infinite series of definitional clauses down into a very simple and finite form. The same is true for our recursive definition of sentences of $\mathcal{L}_S$.

Just as the definition tells us how complex sentences can be built up from simpler parts, the definition tells us how to decompose sentences into their simpler parts. Let's consider some examples. Suppose we want to know whether or not the expression $\neg\neg\neg D$ counts as a sentence of $\mathcal{L}_S$. Looking at the second clause of the definition, we know that $\neg\neg\neg D$ is a sentence only if $\neg\neg D$ is a sentence. So now we need to ask whether or not $\neg\neg D$ is a sentence. Again looking at the second clause of the definition, $\neg\neg D$ is a sentence only if $\neg D$ is. Again, $\neg D$ is a sentence only if D is a sentence. Now, finally, we know that D is an atomic sentence in $\mathcal{L}_S$, so we know that D is a sentence by the first clause of the definition. Since we have systematically decomposed the initial expression $\neg\neg\neg D$ down into its atomic parts according to the rules specified by the definition of a sentence, we know that $\neg\neg\neg D$ counts as a sentence. After all, if we can decompose it in this manner, then we can build it back up—and if we can build the expression $\neg\neg\neg D$ up from just an atomic sentence D plus the rules for constructing new sentences using $\neg$, then $\neg\neg\neg D$ is a sentence. By contrast, you cannot build $\neg\neg D\neg$ up in the same kind of way, so $\neg\neg D\neg$ is a mere expression and not a sentence.

Next, consider the more complex example $\neg(P \wedge \neg(\neg Q \vee R))$, and let's see how this can be "built up" from the atomic sentences P and Q and R:

i) According to clause 1 of the definition, P and Q and R are all sentences.
ii) According to clause 2, if Q is a sentence then $\neg Q$ is a sentence.
iii) According to clause 4, if $\neg Q$ and R are sentences then $(\neg Q \vee R)$ is a sentence.
iv) According to clause 2, if $(\neg Q \vee R)$ is a sentence then $\neg(\neg Q \vee R)$ is a sentence.
v) According to clause 3, if P and $\neg(\neg Q \vee R)$ are sentences then $(P \wedge \neg(\neg Q \vee R))$ is a sentence.
vi) According to clause 2, if $(P \wedge \neg(\neg Q \vee R))$ is a sentence then $\neg(P \wedge \neg(\neg Q \vee R))$ is a sentence.

The definition thus implies that $\neg(P \wedge \neg(\neg Q \vee R))$ counts as a sentence of $\mathcal{L}_S$. For any expression that *cannot* be constructed by the repeated application of these rules, that expression will not count as a sentence of $\mathcal{L}_S$—it is meaningless gibberish. Ultimately, every sentence of $\mathcal{L}_S$ either is an atomic sentence, or it is constructed out of atomic sentences and connectives. The recursive structure of sentences in $\mathcal{L}_S$ will be important in the next chapter, when we consider the circumstances under which sentences of $\mathcal{L}_S$ are true or false.

Bracketing conventions

Strictly speaking, the brackets in $(Q \wedge R)$ are a part of the sentence. Part of the reason for this is because we might want to use $(Q \wedge R)$ as a sub-sentence in a more complicated sentence. For example, we might want to negate $(Q \wedge R)$, obtaining $\neg(Q \wedge R)$. If we just had $Q \wedge R$ without the brackets and put a negation in front of it, we would have $\neg Q \wedge R$, which means something quite different than $\neg(Q \wedge R)$. It is most natural to read $\neg Q \wedge R$ as meaning the same thing as $(\neg Q \wedge R)$.

Strictly speaking, then, $Q \wedge R$ is *not* a sentence. It is a mere *expression*. With that said, when working with $\mathcal{L}_S$, it will make our lives much easier if we are sometimes a little less than strict when it comes to writing down sentences. We will allow ourselves to omit the brackets in a sentence if there are no symbols *outside* of those brackets. Thus, for example, we will allow ourselves to write $Q \wedge R$ instead of the sentence $(Q \wedge R)$. On the other hand, we are *not* allowed to replace $\neg(P \vee Q)$ with $\neg P \vee Q$, because in doing so we would change the meaning of the sentence. Likewise, we are not allowed to replace $P \vee (Q \wedge R)$ with $P \vee Q \wedge R$, since that would introduce ambiguity into the symbolisation.

Whenever we want to turn $Q \wedge R$ into a sub-sentence for a larger, more complex sentence, we need to put the brackets back around it. This is important. A key part of learning how to translate ordinary sentences into logical language is the proper placement of brackets, so do try to make sure you understand how these rules work.

Chapter 3: Key Ideas

▷ Sentences of the form 'If $\mathcal{X}$, then $\mathcal{Y}$' are called **conditionals**, and can be symbolised as $\mathcal{X} \rightarrow \mathcal{Y}$. The sentence on the left of the arrow is called the **antecedent**, and the sentence on the right is called the **consequent**. The characteristic truth table for conditional sentences is:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \rightarrow \mathcal{Y}$
T	T	T
T	F	F
F	T	T
F	F	T

▷ A sentence of the form '$\mathcal{X}$ if and only if $\mathcal{Y}$' is called a **biconditional**, and can be symbolised as $\mathcal{X} \leftrightarrow \mathcal{Y}$. The biconditional $\mathcal{X} \leftrightarrow \mathcal{Y}$ is a really just a shorthand way of expressing the conjunction of two conditionals, $(\mathcal{X} \rightarrow \mathcal{Y}) \wedge (\mathcal{Y} \rightarrow \mathcal{X})$. The characteristic truth table for biconditional sentences is:

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \leftrightarrow \mathcal{Y}$
T	T	T
T	F	F
F	T	F
F	F	T

▷ General rules for symbolisation:

- It is not the case that $P \;=\; \neg P$
- Either P, or $Q \;=\; P \vee Q$
- Neither P, nor $Q \;=\; \neg(P \vee Q)$, or $\neg P \wedge \neg Q$
- Both P, and $Q \;=\; P \wedge Q$
- If P, then $Q \;=\; P \rightarrow Q$
- P only if $Q \;=\; P \rightarrow Q$
- P if and only if $Q \;=\; P \leftrightarrow Q$
- Unless P, $Q \;=\; P \vee Q$, or $\neg P \rightarrow Q$, or $\neg Q \rightarrow P$
- P unless $Q \;=\; P \vee Q$, or $\neg P \rightarrow Q$, or $\neg Q \rightarrow P$

▷ A **sentence** of $\mathcal{L}_S$ is always constructed out of atomic sentences (A, B, C, ...) and connectives ($\neg, \wedge, \vee, \rightarrow, \leftrightarrow$). A sequence of these symbols counts as a sentence of $\mathcal{L}_S$ if and only if it satisfies the definition of a sentence in $\mathcal{L}_S$:

Definition of a SENTENCE OF $\mathcal{L}_S$:

1. Every atomic sentence is a sentence.
2. If $\mathcal{X}$ is a sentence, then $\neg\mathcal{X}$ is a sentence.
3. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \wedge \mathcal{Y})$ is a sentence.
4. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \vee \mathcal{Y})$ is a sentence.
5. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \rightarrow \mathcal{Y})$ is a sentence.
6. If $\mathcal{X}$ and $\mathcal{Y}$ are sentences, then $(\mathcal{X} \leftrightarrow \mathcal{Y})$ is a sentence.
7. Nothing else is a sentence.

Practice Exercises

Part A
Using the symbolisation key given, translate each of the following into $\mathcal{L}_S$ as close as possible.

A : Mister Ace was murdered.
B : The butler did it.
C : The cook did it.
D : The Duchess is lying.
E : Mister Edge was murdered.
F : The murder weapon was a frying pan.

1. Either Mister Ace or Mister Edge was murdered.
2. If Mister Ace was murdered, then the cook did it.
3. If Mister Edge was murdered, then the cook did not do it.
4. Either the butler did it, or the Duchess is lying.
5. The cook did it only if the Duchess is lying.
6. If the murder weapon was a frying pan, then the culprit must have been the cook.
7. Mister Ace was murdered if and only if Mister Edge was not murdered.
8. The Duchess is lying, unless it was Mister Edge who was murdered.
9. If Mister Ace was murdered, he was done in with a frying pan.
10. Since the cook did it, the butler did not.
11. Of course the Duchess is lying!

⋆ **Part B**
Using the symbolisation key given, translate the following into its nearest $\mathcal{L}_S$ equivalent.

E_1 : Adam is an entertainer.
E_2 : Heather is an entertainer.
F_1 : Adam is a philosopher.
F_2 : Heather is a philosopher.
S_1 : Adam is satisfied with his career.
S_2 : Heather is satisfied with her career.

1. Adam and Heather are both entertainers.
2. If Adam is a philosopher, then he is satisfied with her career.
3. Adam is a philosopher, unless he is an entertainer.
4. Heather is an unsatisfied entertainer.
5. Neither Adam nor Heather is an entertainer.
6. Both Adam and Heather are entertainers, but neither of them find it satisfying.
7. Heather is satisfied only if she is a philosopher.
8. If Adam is not an entertainer, then neither is Heather, but if he is, then she is too.
9. Adam is satisfied with his career if and only if Heather is not satisfied with hers.
10. If Heather is both an entertainer and a philosopher, then she must be satisfied with her career.
11. It cannot be that Heather is both an entertainer and a philosopher.
12. Heather and Adam are both philosophers if and only if neither of them is an entertainer.

⋆ Part C

Give a symbolisation key and translate the following into its nearest $\mathcal{L}_S$ equivalent.

1. Alice and Bob are both spies.
2. If either Alice or Bob is a spy, then the code has been broken.
3. If neither Alice nor Bob is a spy, then the code remains unbroken.
4. The German embassy will be in an uproar, unless someone has broken the code.
5. Either the code has been broken or it has not, but the German embassy will be in an uproar regardless.
6. Either Alice or Bob is a spy, but not both.

Part D

For each argument, give a symbolisation key and translate the following into its nearest $\mathcal{L}_S$ equivalent.

1. If Dorothy plays the piano in the morning, then Roger wakes up cranky. Dorothy plays piano in the morning unless she is distracted. So if Roger does not wake up cranky, then Dorothy must be distracted.
2. It will either rain or snow on Tuesday. If it rains, Neville will be sad. If it snows, Neville will be cold. Therefore, Neville will either be sad or cold on Tuesday.
3. If Zoog remembered to do his chores, then things are clean but not neat. If he forgot, then things are neat but not clean. Therefore, things are either neat or clean—but not both.

Chapter 4

Truth Tables

This chapter introduces a way of evaluating the logical status of complex sentences and arguments formalised in the language of $\mathcal{L}_S$. Although it can be laborious, the truth table method is a purely mechanical procedure that requires no intuition or special insight. That may sound like a bad thing, but in fact it's incredibly useful: so long as you follow the method correctly, you're guaranteed to get the correct result.

1. Complete truth tables

For reference, the following tables summarise the characteristic truth tables for the five truth-functional connectives outlined in the previous two chapters.

$\mathcal{X}$	$\neg\mathcal{X}$
T	F
F	T

$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \wedge \mathcal{Y}$	$\mathcal{X} \vee \mathcal{Y}$	$\mathcal{X} \to \mathcal{Y}$	$\mathcal{X} \leftrightarrow \mathcal{Y}$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	F	T	T	F
F	F	F	F	T	T

Keep in mind that $\mathcal{X}$ and $\mathcal{Y}$ can represent *any* sentences in $\mathcal{L}_S$ whatsoever—so not *only* atomic sentences. For example, in §2.1 we said that 'The widget is irreplaceable' should be symbolised as $\neg R$ and 'The widget is not irreplaceable' as $\neg\neg R$. The truth table for negation tells us how the truth-value for $\neg R$ depends on the truth-value for its negand R:

R	$\neg R$
T	F
F	T

But the table *also* tells us how the truth-value of $\neg\neg R$ depends on the value of its negand $\neg R$. To see this, suppose that we use the special symbol $\mathcal{R}$ to represent the non-atomic sentence $\neg R$; so, writing $\mathcal{R}$ is just another way of writing $\neg R$. Then, $\neg\neg R$ just means the same thing as $\neg\mathcal{R}$. We can therefore apply the characteristic truth table for negation to $\neg\mathcal{R}$:

$\mathcal{R}$	$\neg\mathcal{R}$
T	F
F	T

$\neg\neg R$ is related to $\neg R$ in just the same way that $\neg R$ is related to R; that is, $\neg\neg R$ is true whenever $\neg R$ is false, and false whenever $\neg R$ is true. And we could extend this reasoning for indefinitely many iterations of the negation operation. For example, $\neg\neg\neg R$ is just the negation of $\neg\neg R$, and therefore has the opposite truth-value of its negand $\neg\neg R$. And $\neg\neg\neg\neg R$ is just the negation of $\neg\neg\neg R$, and so on, *ad infinitum*. This means that we can use truth tables to work out how the truth-values of even very complex non-atomic sentences ultimately depend on the truth-values of their atomic parts. To get the feel for how we can do this, let's look at a slightly more complicated example.

Complete truth tables

Suppose we have a non-atomic sentence $(H \wedge I) \rightarrow H$. What we want to do is work out the conditions (if any) under which $(H \wedge I) \rightarrow H$ will be true, and when it will be false.

The first step is to mark down all possible combinations of truth and falsity for H and I. As there are only two atomic sentences here, there are *four* possible combinations and therefore four rows on our truth table, like so:

H	I	$(H \wedge I) \rightarrow H$
T	T	
T	F	
F	T	
F	F	

We then copy the truth-values for H and I over and write them underneath the relevant letters on the right-hand side of the table:

H	I	$(H$	$\wedge$	$I\,)$	$\rightarrow$	H
T	T	**T**		**T**		**T**
T	F	**T**		**F**		**T**
F	T	**F**		**T**		**F**
F	F	**F**		**F**		**F**

Now, the sentence $(H \wedge I) \rightarrow H$ *as a whole* is a conditional. That is, $(H \wedge I) \rightarrow H$ has the form $\mathcal{X} \rightarrow \mathcal{Y}$, where

$$\mathcal{X} = H \wedge I$$
$$\mathcal{Y} = H$$

Since the antecedent is itself a complex sentence with the form of a conjunction, we can break it down into its parts as well. So we will begin by filling in the truth table just for the conjunctive antecedent.

H and I are both true on the first row. The first row represents the possible situation where H and I are both true. Since the characteristic truth table for conjunction tells us that a conjunction is true when both conjuncts are true, we write T underneath the conjunction symbol for the first row:

H	I	$(H$	$\wedge$	$I\,)$	$\rightarrow$	H
		$\mathcal{X}$	$\wedge$	$\mathcal{Y}$		
T	T	T	**T**	T		T
T	F	T		F		T
F	T	F		T		F
F	F	F		F		F

The second row represents the possible situation where H is true and I is false. So, since a conjunction is false whenever one of its conjuncts is false, we write F underneath the conjunction symbol for the second row:

H	I	$(H$	$\land$	$I)$	$\to$	H
		$\mathcal{X}$	$\land$	$\mathcal{Y}$		
T	T	T	**T**	T		T
T	F	T	**F**	F		T
F	T	F		T		F
F	F	F		F		F

We continue this process for final two rows and get this:

H	I	$(H$	$\land$	$I)$	$\to$	H
		$\mathcal{X}$	$\land$	$\mathcal{Y}$		
T	T	T	**T**	T		T
T	F	T	**F**	F		T
F	T	F	**F**	T		F
F	F	F	**F**	F		F

The column under the conjunction symbol $\land$ thus tells us that the antecedent, $(H \land I)$, is true when H and I are true, and false otherwise.

Almost done! As we noted just above, the entire sentence is a conditional: it has the form $\mathcal{X} \to \mathcal{Y}$. And, as the characteristic truth table for the material conditional then tells us, $(H \land I) \to H$ is false just in case $(H \land I)$ is true and H is false; it is true in all other cases. So now we're going to use what we've just worked out about the truth-value of the antecedent $(H \land I)$ to determine the truth-value of the whole sentence $(H \land I) \to H$.

For the first row, we've worked that $(H \land I)$ is true. H is also true on that row. Therefore, the whole conditional must be true on that row, and we put down a T for the first row underneath the $\to$ symbol. This tells us that $(H \land I) \to H$ is true if H and I are both true:

H	I	$(H \land I)$	$\to$	H
		$\mathcal{X}$	$\to$	$\mathcal{Y}$
T	T	T	**T**	T
T	F	F		T
F	T	F		F
F	F	F		F

Then, on the second row, $(H \land I)$ is false, and H is true. Since a conditional is true whenever the antecedent is false, we can again write a T in the second row underneath the conditional symbol:

H	I	$(H \land I)$	$\to$	H
		$\mathcal{X}$	$\to$	$\mathcal{Y}$
T	T	T	**T**	T
T	F	F	**T**	T
F	T	F		F
F	F	F		F

We continue for the final two rows and get this:

H	I	$(H \wedge I)$	$\to$	H
		$\mathcal{X}$	$\to$	$\mathcal{Y}$
T	T	T	**T**	T
T	F	F	**T**	T
F	T	F	**T**	F
F	F	F	**T**	F

Now the column underneath the conditional tells us that $(H \wedge I) \to I$ is always true, regardless of the truth-values of H and I. This should be intuitive: if H *and* I are true, then H by itself is true. We can know that this must be the case purely by our logical reasoning; the truth table confirms it.

In this example, we have not repeated all of the entries in every successive table. When actually writing truth tables on paper, however, it is impractical to erase whole columns or rewrite the whole table for every step. Although it is more crowded, the complete truth table can be written in this way:

H	I	$(H$	$\wedge$	$I)$	$\to$	H
T	T	T	T	T	**T**	T
T	F	T	F	F	**T**	T
F	T	F	F	T	**T**	F
F	F	F	F	F	**T**	F

The truth-value of the sentence on each row is just the column underneath the **main logical connective** of the sentence; in this case, that will be the column underneath the conditional $\to$. Each row of the table represents whether $(H \wedge I) \to I$ at a different possibility. The first row represents the possibility where H and I are both true, the next two rows represent possibilities where one is true but not the other, and the final row represents the possibility where both H and I are false.

The main logical connective

Working out which connective is the main connective in a complex sentence can sometimes be quite difficult. The main connective is defined as the connective with the **widest scope**. The relative scope of a connective within a sentence corresponds to the order in which it's added when constructing the sentence using the rules defined for sentence construction we looked at in the previous chapter. We will begin with a simple example:

$$(A \leftrightarrow B) \to B$$

If you were to construct this sentence out of its atomic parts using the rules for sentence construction, it would go like this:

1. A and B are sentences
2. Since A and B are sentences, $(A \leftrightarrow B)$ is a sentence
3. Since $(A \leftrightarrow B)$ and B are sentences, $(A \leftrightarrow B) \to B$ is a sentence

In general, the main connective is always the connective that's added at the final step in the construction of the sentence. The final step of the foregoing construction involves connecting $(A \leftrightarrow B)$ and B together using $\rightarrow$, so $\rightarrow$ is the main connective. It has the widest scope, because it connects $(A \leftrightarrow B)$ on its left side and B on its right. In this case, the $\leftrightarrow$ is included **within** the scope of the $\rightarrow$, and therefore we say the $\leftrightarrow$ has narrower scope than $\rightarrow$.

Now consider a more complicated example:

$$((A \leftrightarrow B) \rightarrow C) \wedge \neg(\neg D \rightarrow C)$$

In this case, the main connective is the conjunction $\wedge$. We can see this, again, by considering how the sentence would be constructed. We already know how the left conjunct $((A \leftrightarrow B) \rightarrow C)$ is constructed, so here are the remaining steps:

1. C and D are sentences
2. Since D is a sentence, $\neg D$ is a sentence
3. Since $\neg D$ and C are sentences, $(\neg D \rightarrow C)$ is a sentence
4. Since $(\neg D \rightarrow C)$ is a sentence, $\neg(\neg D \rightarrow C)$ is a sentence
5. (from earlier:) $((A \leftrightarrow B) \rightarrow C)$ is a sentence
6. Since $((A \leftrightarrow B) \rightarrow C)$ and $\neg(\neg D \rightarrow C)$ are sentences, $((A \leftrightarrow B) \rightarrow C) \wedge \neg(\neg D \rightarrow C)$ is a sentence

The final step of the construction is connecting $((A \leftrightarrow B) \rightarrow C)$ and $\neg(\neg D \rightarrow C)$ together using $\wedge$. So the sentence $((A \leftrightarrow B) \rightarrow C) \wedge \neg(\neg D \rightarrow C)$, as a whole, is a conjunction, and $\wedge$ is the main connective. The $\wedge$ includes $((A \leftrightarrow B) \rightarrow C)$ within its scope on the left side, so the $\rightarrow$ and $\leftrightarrow$ on the left have a narrower scope than $\wedge$. Likewise, the $\wedge$ includes $\neg(\neg D \rightarrow C)$ on the right side, so the $\rightarrow$ and both of the $\neg$ symbols on the right have narrower scope than $\wedge$.

What about $\neg(\neg D \rightarrow C)$? In this case, the connective with the widest scope is the left-most negation; next is the $\rightarrow$, and then finally the second $\neg$ has the narrowest scope. This is apparent in how the sentence as a whole is constructed out of D and C, plus the rules for $\neg$ and $\rightarrow$ (steps 1 through 4 of the foregoing construction).

When we're filling in truth tables for a complex sentence, we always begin with the connectives that have the narrowest scope, and work out way up to the main logical connective. This will correspond to the order in which the sentence is constructed out of its atomic parts. The connectives with the *narrowest* scope always connect to simple sentence letters—there should be no complex sentences within their scope. In the end, there should only ever be *one* main logical connective.

The main connective tells you what *kind* of sentence you're dealing with. So, since the main connective of $((A \leftrightarrow B) \rightarrow C) \wedge \neg(\neg D \rightarrow C)$ is a $\wedge$, the sentence as a whole is a conjunction; and since the main connective of $(H \wedge I) \rightarrow I$ earlier was $\rightarrow$, that sentence is a conditional.

> The MAIN LOGICAL CONNECTIVE of any complex sentence is the connective with the widest scope. There is only ever one main logical connective. The column of the truth table that is underneath the main logical connective indicates the conditions under which the sentence is true.

Truth table rows

A complete truth table for any sentence should have a row for all the possible combinations of T and F, for all of the sentence letters that can be found in the symbolization. The size of the complete truth table therefore depends on the number of different sentence letters in the table. This is true even if the same letter is repeated many times, as in $((C \leftrightarrow C) \rightarrow C) \wedge \neg(\neg C \rightarrow C)$. The complete truth table for this requires only two lines because there are only two possibilities: C can be true, or it can be false. The truth table for this sentence looks like this:

C	$((C \leftrightarrow C) \rightarrow C) \wedge \neg(\neg C \rightarrow C)$
T	T T T T T **F** F F T T T
F	F T F F F **F** T T F F F

Note that a single sentence letter can never be marked both T and F *on the same row.* Each row represents a possible truth-value for C.

A complex sentence that contains two atomic sentences as parts requires *four* lines for a complete truth table, as in the characteristic truth tables and the table for $(H \wedge I) \rightarrow I$. Here there are four possible combinations: TT, TF, FT, and FF. A complex sentence that contains three atomic sentences as parts requires *eight* lines. For example:

M	N	P	$M \wedge (N \vee P)$
T	T	T	T **T** T T T
T	T	F	T **T** T T F
T	F	T	T **T** F T T
T	F	F	T **F** F F F
F	T	T	F **F** T T T
F	T	F	F **F** T T F
F	F	T	F **F** F T T
F	F	F	F **F** F F F

From this table, we know that $M \wedge (N \vee P)$ might be true or false, depending on the truth-values of M, N, and P. In particular, the truth table tells us that $M \wedge (N \vee P)$ is true in exactly three cases:

1. When M, N, and P are all true.
2. When M and N are true, and P false.
3. When M and P are true, and N false.

This holds *regardless* of what the sentence letters M, N, and P refer to. This is because the connectives $\wedge$ and $\vee$ are truth-functional: the truth of conjunction/disjunction depends only on the truth of the conjuncts/disjuncts. Thus we can break down complex sentences into their smaller parts, work out the truth-values of those parts, and work back up to the truth-value of the whole.

A complete truth table for a complex sentence that contains four atomic sentences as parts requires 16 lines. Five atomic sentences as parts, 32 lines. Six atomic sentences as parts, 64 lines. And so on. To be perfectly general: if a complex sentence has n-many atomic sentences as parts, then its complete truth table must have 2^n rows. So if a complex sentence has 10 atomic sentences as its parts, its truth table must have $2^{10} = 1024$ rows. (Don't worry, we won't consider any sentences with 10 atomic sub-sentences.)

In order to fill in the columns of a complete truth table, begin with the right-most sentence letter and alternate Ts and Fs. In the next column to the left, write two Ts, write two Fs, and repeat. For the third sentence letter, write four Ts followed by four Fs. This yields an eight line truth table like the one above. For a 16 line truth table, the next column of sentence letters should have eight Ts followed by eight Fs. And so on.

2. Tautologies and Contradictions

Using truth tables, we can quickly come to see that there may be some sentences that are *always* and *necessarily* true just by virtue of their logical form. That is, there may be some complex sentences that are always true regardless of the truth-values that their atomic parts take. We have already seen some examples of this in the previous section, but here is a rather simple example:

1. Either dinosaurs lived on Earth, or they didn't.

Let's symbolise this and fill in its truth table. We'll use the following symbolisation key:

D : Dinosaurs lived on Earth.

With this key, we should symbolise Sentence 1 as $D \vee \neg D$. Now when we fill in its truth table, we'll see that it must be true regardless of whether D is true or false:

D	D	$\vee$	$\neg$	D
T	T	**T**	F	T
F	F	**T**	T	F

This should be intuitive: regardless of what happened in the past, either dinosaurs lived on Earth, or they didn't. One of these two disjuncts must be true. We can know this without having to go out and check; merely doing some logical reasoning is enough. Indeed, the same can be said of any sentence with the same logical form. For instance:

2. The dinosaurs were wiped out by an asteroid, or they weren't.
3. I am human, or non-human.
4. Whales are mammals, or they are not.

The specific content of these sentences does not matter so much for working out their truth-value—what matters is that they have a logical form which forces them to be true. Their truth tables reflect this fact, in that they have all Ts underneath their main connective. We will call any sentence which, once it has been symbolised in $\mathcal{L}_S$, has all Ts underneath its main connective, a **tautology** in $\mathcal{L}_S$.

(For simplicity, in the discussion that follows we'll usually leave out the 'in $\mathcal{L}_S$' clause, but it's important to keep it in the back of your mind. Whether a sentence of English has a tautologous logical form depends in part upon how *we* formalise it—that is, it depends on the formal language we use when symbolising it. As we will see later chapters, there are some sentences which seem to be true purely as a matter of logic, but which *aren't* tautologies **in** $\mathcal{L}_S$. However, they might still be tautologies when formulated in a more powerful formal language such as the one we will develop in the next chapter.)

Another simple tautology can be seen from the following truth table:

D	$D \rightarrow D$
T	T **T** T
F	F **T** F

Again, this should be very intuitive: *if* dinosaurs lived on Earth, *then* dinosaurs lived on Earth. Nothing could be more obviously true. Propositions with this form are also true as a matter of logic, they cannot be false. One more example before we move on:

D	$\neg(D \wedge \neg D)$
T	**T** T F F T
F	**T** F F T F

In this case we have '*It is not the case that* dinosaurs lived on Earth and they didn't live on Earth'. There are in fact infinitely many tautologies in $\mathcal{L}_S$, and many of them are very, very complex.

Notice that the truth-values underneath the $\wedge$ in $\neg(D \wedge \neg D)$ are always false. So if we were to fill in the truth table for $D \wedge \neg D$, we'd find that there are only Fs underneath its main connective. We will say that any sentence which has this property is a **contradiction** in $\mathcal{L}_S$. As with tautologies, every contradiction in $\mathcal{L}_S$ will also be a contradiction in $\mathcal{L}_P$. So, if something is a contradiction in $\mathcal{L}_S$, we can know that it must be false merely using logic. But there will also be some contradictions in the more powerful logical language $\mathcal{L}_P$ which are *not* contradictions in $\mathcal{L}_S$.

Contradictions are the opposite of tautologies: they have such a form that they *cannot* be true, regardless of how the world turns out to be. $\neg(D \wedge \neg D)$ is a tautology precisely because it says of some contradiction—*viz.*, $D \wedge \neg D$—that it is not true. Since contradictions are always false, the negation of a contradiction is always true. Likewise, the negation of every tautology is a contradiction.

> A sentence of $\mathcal{L}_S$ is a TAUTOLOGY if its truth table shows all Ts underneath its main logical connective. A sentence of $\mathcal{L}_S$ is a CONTRADICTION if its truth table shows all Fs underneath its main logical connective.

We can say that a sentence is **contingent** in $\mathcal{L}_S$ if its truth table shows some Ts and some Fs underneath its main connective. If we are dealing with a contingent sentence, then for all the truth table method tells us, the sentence may be true or it may be false. Finally, we can say that a sentence is **consistent** in $\mathcal{L}_S$ just in case its truth table has at least one T underneath its main connective. Thus, a consistent sentence can be either a tautology or contingent, but not a contradiction.

> A sentence of $\mathcal{L}_S$ is CONTINGENT if its truth table shows at least one T and at least one F underneath its main logical connective. A sentence of $\mathcal{L}_S$ is CONSISTENT if its truth table shows at least one T under its main logical connective.

Every sentence in $\mathcal{L}_S$ is either consistent or a contradiction. If it is consistent, then it is either contingent or a tautology, and if it is either contingent or a tautology, then it is consistent.

3. Logical Equivalence

We can also ask about the logical relationships *between* collections of sentences. We will begin with what it means for two (or more) sentences to be **logically equivalent** when symbolized in $\mathcal{L}_S$. For example:

J_1 : John went to the store and to the park.
J_2 : John went to the park and to the store.

J_1 and J_2 are both contingent, since John might not have gone anywhere at all. Yet they must have the same truth-value. If either is true, then they both are. And if either is false, then they both are. When two sentences have the same truth values under all the same conditions, we say that they are logically equivalent.

Consider the sentences $\neg(A \vee B)$ and $\neg A \wedge \neg B$. Are they logically equivalent? We can check by constructing a **combined truth table** for both sentences at once:

A	B	$\neg(A \vee B)$	$\neg A \wedge \neg B$
T	T	**F** TTT	FT**F**FT
T	F	**F** TTF	FT**F**TF
F	T	**F** FTT	TF**F**FT
F	F	**T** FFF	TF**T**TF

Look at the columns for the main connectives; negation for the first sentence, and conjunction for the second. On the first three rows, both are F. On the final row, both are T. They have the same pattern of Ts and Fs under their main connectives. Since they match on every row, the two sentences are logically equivalent—this means that they are true or false under exactly the same conditions. In other words, $\neg(A \vee B)$ is true if and only if $\neg A \wedge \neg B$ is true.

As it turns out, there are many interesting logical equivalencies. For example, there is also the equivalence of $\neg(A \wedge B)$ and $\neg A \vee \neg B$. Another example is that $A \rightarrow B$ is equivalent to $\neg(A \wedge \neg B)$, which is equivalent to $\neg A \vee B$, which is equivalent to $\neg B \rightarrow \neg A$. In fact, for any sentence of $\mathcal{L}_S$, there are infinitely many logically equivalent sentences.

As we said in §3.2, a biconditional is true just in case the left-hand side of the biconditional always has the same truth-value as the right-hand side. Therefore, if we have two logically equivalent sentences—for example, $\neg(A \vee B)$ and $\neg A \wedge \neg B$—we can put them on either side of a biconditional to create a tautology:

A	B	$\neg(A \vee B) \leftrightarrow (\neg A \wedge \neg B)$
T	T	F TTT **T** FTFFT
T	F	F TTF **T** FTFTF
F	T	F FTT **T** TFFFT
F	F	T FFF **T** TFTTF

So, every *pair* of logically equivalent sentences corresponds to a tautologous biconditional (and *vice versa*). In other words, if a biconditional sentence is a tautology, then the left-hand side and the right-hand side of the biconditional are logically equivalent. And, in the other direction: if two sentences are logically equivalent, then the biconditional between them is a tautology. This means we can use the truth tables of biconditionals to test for logical equivalence.

Two sentences of $\mathcal{L}_S$ are LOGICALLY EQUIVALENT if they have the same truth table—that is, if they have the same pattern of Ts and Fs under their main connective. Alternatively, two sentences are logically equivalent if the biconditional between them is a tautology.

4. Joint Consistency

Next consider these two sentences:

B_1 : Bob went for a swim today.
B_2 : Bob did not go for a swim today.

Logic alone cannot tell us which, if either, of these two sentences is true. Each sentence is consistent when considered in isolation from the other. Yet we can say that *if* the first (B_1) is true, *then* the second (B_2) must be false. Likewise, if the second sentence is true, then the first sentence must be false. It cannot be the case that both of these sentences are true at the same time.

If the members of a set of sentences cannot all be true at the same time in this kind of way, then the *set* is said to be **inconsistent**. This is different from saying that each sentence within the set is inconsistent—it might be the case that a set of sentences is inconsistent even though every sentence within the set is consistent. If we have an inconsistent set of sentences, then we can also say that the sentences within the set are **jointly inconsistent**. To say that a set of sentences is inconsistent, and to say that the sentences within the set are jointly inconsistent, are just different ways of saying the same thing.

We can ask about the consistency of any set of sentences. For example, consider the following sentences:

G_1 : There are at least four giraffes at the wild animal park.
G_2 : There are exactly seven gorillas at the wild animal park.
G_3 : There are not more than two Martians at the wild animal park.
G_4 : Every giraffe at the wild animal park is a martian.

G_1 and G_4 together imply that there are at least four martians at the park. Together they are consistent, but they also conflict with G_3, which implies that there are no more than Martians at the park. So the set of sentences that includes all of the sentence G_1 through to G_4—i.e., the set $\{G_1, G_2, G_3, G_4\}$—is inconsistent, even though every sentence in the set is consistent.

(The curly brackets {, }, are used in this case to represent sets of sentences. So $\{A, B\}$ is the set that includes the sentences A and B, and nothing else, while $\{A, B, C\}$ is the set that includes sentences A, B, and C, and nothing else.)

Notice that the inconsistency of the set $\{G_1, G_2, G_3, G_4\}$ has nothing at all to do with G_2—it just happens to be a member of an inconsistent set of sentences, but its membership does not contribute to that inconsistency. If a set of sentences is inconsistent, *adding* more sentences to the set will also result in an inconsistent set of sentences. In this case, the set $\{G_1, G_3, G_4\}$ is inconsistent because of the conflict between G_1, G_3, and G_4, and therefore any larger set such as $\{G_1, G_2, G_3, G_4\}$ must also be inconsistent. You cannot remove inconsistency by adding sentences. On the other hand, if a set of sentences is consistent, adding new sentences may create an inconsistency. The set $\{G_1, G_4\}$ is consistent, but $\{G_1, G_3, G_4,\}$ is not. So you can remove consistency by adding sentences.

We can use truth tables to test for the consistency of any set of sentences. Suppose we have three sentences, $P \wedge Q$, $P \to R$, and $R \to \neg Q$. Individually, the truth tables for these three are:

P	Q	$P \wedge Q$
T	T	T **T** T
T	F	T **F** F
F	T	F **F** T
F	F	F **F** F

P	R	$P \to R$
T	T	T **T** T
T	F	T **F** F
F	T	F **T** T
F	F	F **T** F

R	Q	$R \to \neg Q$
T	T	T **F** F T
T	F	T **T** T F
F	T	F **T** F T
F	F	F **T** T F

As you can see, each sentence in the set is consistent. To work out whether they're jointly consistent we'll need to put all three into a single combined truth table:

P	Q	R	$P \wedge Q$	$P \to R$	$R \to \neg Q$
T	T	T	T **T** T	T **T** T	T **F** F T
T	T	F	T **T** T	T **F** F	F **T** F T
T	F	T	T **F** F	T **T** T	T **T** T F
T	F	F	T **F** F	T **F** F	F **T** T F
F	T	T	F **F** T	F **T** T	T **F** F T
F	T	F	F **F** T	F **T** F	F **T** F T
F	F	T	F **F** F	F **T** T	T **T** T F
F	F	F	F **F** F	F **T** F	F **T** T F

Now, if we go through each row, we'll see that there is no row where all three of the sentences are true. That means that there is no possible condition—no combination of truth-values for P, Q, and R—under which all three of $P \wedge Q$, $P \to R$, and $R \to \neg Q$ can be true at the same time. So those three sentences are jointly inconsistent.

Interestingly, if some sentences are jointly inconsistent, then the *conjunction* of all those sentences together will always be a contradiction. We can see this with the following truth table for the conjunction of $P \wedge Q$, $P \to R$, and $R \to \neg Q$:

P	Q	R	$(P \wedge Q) \wedge ((P \to R) \wedge (R \to \neg Q))$
T	T	T	T T T **F** T T T F T F F T
T	T	F	T T T **F** T F F F F T F T
T	F	T	T F F **F** T T T F T F F T
T	F	F	T F F **F** T F F F F T F T
F	T	T	F F T **F** F T T T T T T F
F	T	F	F F T **F** F T F T F T T F
F	F	T	F F F **F** F T T T T T T F
F	F	F	F F F **F** F T F T F T T F

So, every inconsistent set of sentences corresponds to a contradictory conjunction (and *vice versa*). This makes sense: a conjunction is only true when each of the conjuncts are true, and a set of sentences is only consistent when each member of the set *can* be true at the same time.

A set of sentences of $\mathcal{L}_S$ is INCONSISTENT if there is no row in the combined truth table under which each sentence is true, or if the conjunction of all those sentences is a contradiction. If a set of sentences is inconsistent, then we say the sentences within the set are JOINTLY INCONSISTENT.

5. Truth Tables and Validity

As we've defined it in Chapter 1, an argument is valid just in case it's not possible its premises to be true and its conclusion false. And, just as we've used truth tables to test for consistency, equivalence, and so on, we can also use truth tables to test for formal validity.

Consider this argument:

P1 $\neg L \to (J \vee L)$
P2 $\neg L$

C J

Is it valid? There are two ways to test this using truth tables. The easiest is to construct a combined truth table that contains each of the premises as well as the conclusion. If the conclusion is true whenever the premises are all true, then the argument is valid.

J	L	$\neg L \to (J \vee L)$	$\neg L$	J
T	T	F T **T** T T T	**F** T	**T**
T	F	T F **T** T T F	**T** F	**T**
F	T	F T **T** F T T	**F** T	**F**
F	F	T F **F** F F F	**T** F	**F**

There is exactly one condition under which both of the premises are true, which is represented by the second row of the truth table, and on that row the conclusion is also true. The truth table is therefore telling us that if all the premises are true then the conclusion must also be true—so the argument is valid.

Note that we *don't* have to worry about the rows where one or more of the premises is false. This means that if there is *no* row where all of the premises are true, then the argument is *automatically* valid. You might find this strange, but it fits with the definition of validity: that it is not possible for the premises to all be true and the conclusion false. If it's not possible for the premises to all be true, then the definition is trivially satisfied, regardless of the truth-value of the conclusion. Hence, arguments with jointly inconsistent premises are always valid—but they will never be sound.

Another slightly strange consequence of the definition is that any argument with a tautologous conclusion is automatically valid—regardless of what its premises are! Again, this fits with the definition of validity: if the conclusion is necessarily true, then it cannot be false; therefore, it cannot be the case that the premises are true and the conclusion false.

An argument is FORMALLY VALID IN $\mathcal{L}_S$ whenever there is no row in its combined truth table such that the premises are all true and the conclusion false. An argument is automatically valid whenever either it has jointly inconsistent premises, or it has a tautologous conclusion.

We can also test for validity by turning the argument into a material conditional. First, we conjoin each of the premises. This conjunction will form the antecedent. Then, the conclusion will be our consequent. If the conditional is a tautology, then the argument is valid. So, for the current example, we begin by conjoining our two premises to get $(\neg L \to (J \vee L)) \wedge \neg L$. This will be the antecedent. The conclusion is just J. Hence, we have $((\neg L \to (J \vee L)) \wedge \neg L) \to J$, the truth table of which is:

J	L	$((\neg L \to (J \lor L)) \land \neg L) \to J$
T	T	F T T T T T F F T **T** T
T	F	T F T T T F T T F **T** T
F	T	F T T F T T F F T **T** F
F	F	T F F F F F F T F **T** F

In general, every argument that is valid when symbolised in $\mathcal{L}_S$ corresponds to material conditional that is a tautology in $\mathcal{L}_S$ (and *vice versa*). This makes sense. The material conditional says that the antecedent is sufficient for the truth of the consequent, and a material conditional is a tautology just in case its antecedent is sufficient for its consequent by virtue of their respective logical forms. On the other hand, a valid argument is one where the premises are jointly sufficient for the truth of the conclusion, by virtue of logical form. Hence, the *conjunction* of the premises is sufficient for the truth of the conclusion, by virtue of logical form.

> An argument with premises P1, P2, …, P$_n$ and conclusion C is FORMALLY VALID IN $\mathcal{L}_S$ whenever the conditional which has (a) the conjunction of P1, P2, …, P$_n$ as its antecedent, and (b) C as its consequent, is a tautology.

Chapter 4: Key Ideas

- ▷ The characteristic truth tables for the different connectives can be applied to specific sentences to determine the conditions under which that sentence is true or false. For example, we apply the characteristic truth tables for ¬ and ∧ to determine the truth conditions for the sentence $A \wedge \neg B$, like so:

A	B	$A \wedge \neg$ B
T	T	T **F** F T
T	F	T **T** T F
F	T	F **F** F T
F	F	F **F** T F

  The column under the main connective (∧) tells us that $A \wedge \neg B$ is true just when A is true and B is false; in all other cases $A \wedge \neg B$ is false.

- ▷ The **main connective** of any complex sentence is the connective with the widest scope. There is only ever one main logical connective. The column of the truth table that is underneath the main logical connective indicates the conditions under which the sentence is true.

- ▷ A sentence is a **tautology** if it is true in any circumstance. This will be so if its truth table has all Ts under its main connective. A sentence is a **contradiction** if it is false in any circumstance. This will be so if its truth table has all Fs under its main connective.

- ▷ A sentence is **contingent** if it is true in some circumstances but not others. This will be so if its truth table has some Ts and some Fs under its main connective. A sentence is **consistent** if it is true in at least one circumstance. This will be so if its truth table has at least one T under its main connective.

- ▷ A set of sentences is **consistent** if there is a possible circumstance where they are all true. If a set of sentences is consistent, then the conjunction of all sentences in the set will be consistent. A set of sentences is **inconsistent** if there is no possible circumstance where they are all true. If a set of sentences is inconsistent, then their conjunction will be inconsistent. Thus we can use truth tables to test for the consistency/inconsistency of a set of sentences.

- ▷ Two sentences are **equivalent** when they are true at exactly the same circumstances. If two sentences $\mathcal{X}$ and $\mathcal{Y}$ are equivalent, then the biconditional $\mathcal{X} \leftrightarrow \mathcal{Y}$ will be a tautology. Thus we can use truth tables to test for equivalence in $\mathcal{L}_S$.

- ▷ An argument is valid just in case, if all the premises are true, then the conclusion must also be true. If an argument is valid, then the conditional sentence which has (a) the conjunction of the premises as its antecedent, and (b) the conclusion as its consequent, will be a tautology. Thus we can use truth tables to test for the validity of an argument.

Practice Exercises

Part A
Determine the main connective for each of the following sentences:

1. $\neg(A \vee B)$
2. $\neg(A \wedge \neg A)$
3. $\neg\neg\neg B$
4. $\neg A \vee B$
5. $\neg\neg A \rightarrow (A \vee \neg B)$
6. $\neg((A \wedge B) \leftrightarrow A)$
7. $((A \wedge B) \wedge \neg(A \wedge B)) \wedge C$

⋆ Part B
Determine whether each sentence is a tautology, a contradiction, or a contingent sentence (in $\mathcal{L}_S$). Justify your answer with a truth table where appropriate.

1. $A \rightarrow A$
2. $\neg B \wedge B$
3. $C \rightarrow \neg C$
4. $\neg D \vee D$
5. $(A \leftrightarrow B) \leftrightarrow \neg(A \leftrightarrow \neg B)$
6. $(A \wedge B) \vee (B \wedge A)$
7. $(A \rightarrow B) \vee (B \rightarrow A)$
8. $\neg(A \rightarrow (B \rightarrow A))$
9. $(A \wedge B) \rightarrow (B \vee A)$
10. $A \leftrightarrow (A \rightarrow (B \wedge \neg B))$
11. $\neg((A \vee B) \leftrightarrow (\neg A \wedge \neg B))$
12. $\neg((A \wedge B) \leftrightarrow A)$
13. $((A \wedge B) \wedge \neg(A \wedge B)) \wedge C$
14. $A \rightarrow (B \vee C)$
15. $((A \wedge B) \wedge C) \rightarrow B$
16. $(A \wedge \neg A) \rightarrow (B \vee C)$
17. $\neg((C \vee A) \vee B$
18. $(B \wedge D) \leftrightarrow (A \leftrightarrow (A \vee C))$

⋆ Part C
Determine whether each pair of sentences below are logically equivalent. Justify your answer with a complete or partial truth table where appropriate.

1. A
   $\neg A$

2. A
   $A \vee A$

3. $A \rightarrow A$
   $A \leftrightarrow A$

4. $A \vee \neg B$
   $A \rightarrow B$

5. $A \wedge \neg A$
   $\neg B \leftrightarrow B$

6. $\neg(A \wedge B)$
   $\neg A \vee \neg B$

7. $\neg(A \rightarrow B)$
   $\neg A \rightarrow \neg B$

8. $A \rightarrow B$
   $\neg B \rightarrow \neg A$

9. $(A \vee B) \vee C$
   $A \vee (B \vee C)$

10. $(A \vee B) \wedge C$
    $A \vee (B \wedge C)$

⋆ Part D
Determine whether each set of sentences is consistent or inconsistent. Justify your answer with a truth table where appropriate.

1. $A \rightarrow A$
   $\neg A \rightarrow \neg A$
   $A \wedge A$
   $A \vee A$

2. $A \wedge B$
   $C \rightarrow \neg B$
   C

3. $A \vee B$
   $A \rightarrow C$
   $B \rightarrow C$

4. $A \rightarrow B$
   $B \rightarrow C$
   A
   $\neg C$

5. $B \wedge (C \vee A)$
   $A \rightarrow B$
   $\neg(B \vee C)$

6. $A \vee B$
   $B \vee C$

$C \to \neg A$

7. $A \leftrightarrow (B \lor C)$
   $C \to \neg A$
   $A \to \neg B$

8. A
   B
   C
   $\neg D$
   $\neg E$
   F

$\star$ **Part E**

Determine whether each argument is valid or invalid. Justify your answer with a truth table where appropriate.

1. $A \to A$
   therefore A

2. $A \lor (A \to (A \leftrightarrow A))$
   therefore A

3. $A \to (A \land \neg A)$
   therefore $\neg A$

4. $A \leftrightarrow \neg(B \leftrightarrow A)$
   therefore A

5. $A \lor (B \to A)$
   therefore $\neg A \to \neg B$

6. $A \to B$
   B
   therefore A

7. $A \lor B$
   $B \lor C$
   $\neg A$
   therefore $B \land C$

8. $A \lor B$
   $B \lor C$
   $\neg B$
   therefore $A \land C$

9. $(B \land A) \to C$
   $(C \land A) \to B$
   therefore $(C \land B) \to A$

10. $A \leftrightarrow B$
    $B \leftrightarrow C$
    therefore $A \leftrightarrow C$

⋆ Part F
Answer each of the questions below and justify your answer.

1. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are logically equivalent. What is the status of $\mathcal{X} \leftrightarrow \mathcal{Y}$? (i.e., is it a tautology, contradiction, contingent?)
2. Suppose that $(\mathcal{X} \wedge \mathcal{Y}) \rightarrow \mathcal{Z}$ is contingent. What can you say about the argument '$\mathcal{X}$, $\mathcal{Y}$, therefore $\mathcal{Z}$'?
3. Suppose that the set of sentences $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$ is inconsistent. What can you say about the status of $(\mathcal{X} \wedge \mathcal{Y}) \wedge \mathcal{Z}$? What about the status of $\mathcal{Y} \wedge (\mathcal{X} \wedge \mathcal{Z})$?
4. Suppose that $\mathcal{X}$ is a contradiction. What can you say about the argument '$\mathcal{X}, \mathcal{Y}$, therefore $\mathcal{Z}$'?
5. Suppose that $\mathcal{Z}$ is a tautology. What can you say about the argument '$\mathcal{X}, \mathcal{Y}$, therefore $\mathcal{Z}$'?
6. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are logically equivalent. What can you say about the status of $\mathcal{X} \vee \mathcal{Y}$?
7. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are *not* logically equivalent. What can you say about the status of $\mathcal{X} \vee \mathcal{Y}$?

Part G
We could leave the symbol for the biconditional $\leftrightarrow$ out of the language if we so desired. Instead of writing $A \leftrightarrow B$ we could just write $(A \rightarrow B) \wedge (B \rightarrow A)$. The fact that those sentences are logically equivalent means that we can *define* the biconditional $\leftrightarrow$ in terms of $\rightarrow$ and $\wedge$ without any loss in our ability to express things. The only loss would be in the simplicity of our symbolisation.

Once we start defining some logical connectives in terms of others, we find out that we actually don't need very many logical connectives at all. Indeed, in many formal languages there are only two connectives. For example, we can define the material conditional in terms of $\neg$ and $\wedge$ as follows:

A	B	A→B	¬(A∧¬B)
T	T	T **T** T	**T** T F F T
T	F	T **F** F	**F** T T T F
F	T	F **T** T	**T** F F F T
F	F	F **T** F	**T** F F T F

We can similarly 'define away' the biconditional and the disjunction without needing anything more than negation and conjunction. Show that this is true by creating sentences equivalent to each of the following without using anything other than negation and conjunction:

⋆ 1. $A \vee B$
⋆ 2. $A \leftrightarrow B$

We could have a language with only negation and disjunction as connectives, but which is just as expressive as $\mathcal{L}_S$. Using negation, and disjunction, work out what sentences are logically equivalent to each of the following:

⋆ 3. $A \to B$
⋆ 4. $A \land B$
⋆ 5. $A \leftrightarrow B$

The *Sheffer stroke*, $|$, is a special two-place logical connective that means 'Not… and…'. (Some people therefore call it 'nand'.) It has the following characteristic truth table:

$\mathcal{X}$	$\mathcal{Y}$	$(\mathcal{X} \mid \mathcal{Y})$
T	T	**F**
T	F	**T**
F	T	**T**
F	F	**T**

Interestingly, *every* sentence written using a connective of $\mathcal{L}_S$ can be rewritten as a logically equivalent sentence using only Sheffer strokes. This means that, strictly speaking, we really only need *one* logical connective in our language without losing any expressive capabilities! (The resulting symbolizations do tend to be very complex, however.)

6. Using only the Sheffer stroke, work out what sentences are logically equivalent to $\neg A$, $A \land B$, $A \lor B$, $A \to B$, and $A \leftrightarrow B$.
7. Do the same, this time using a single symbol for 'Not… or…' instead of the Sheffer stroke.

Chapter 5

Names, Predicates, and Quantifiers

Consider the following argument:

P1 Willard is a logician.
P2 All logicians are pedantic.

C Willard is pedantic.

There is no possible way that both of the premises can be true and the conclusion false, so the argument is valid. And there is something about the *form* of the argument that suggests that it must be valid. However, suppose we try to symbolise it in $\mathcal{L}_S$, using the following symbolisation key:

L : Willard is a logician.
A : All logicians are pedantic.
P : Willard is pedantic.

The choice of the particular letters 'L', 'A' and 'P' is immaterial—what matters is that each of the three sentences does not contain an atomic sentence as a proper part, and so cannot be symbolised in $\mathcal{L}_S$ as complex sentences. The argument would then be represented as follows:

P1 L
P2 A

C P

It's easy to see that this argument doesn't have a valid form when symbolised in $\mathcal{L}_S$. This situation is unfortunate. To be sure, we noted in §1.3 that there are some valid *arguments* with non-valid *argument forms*, so we should sometimes expect that our logic won't capture everything there is to validity. However, it would be nice if there were a way to capture what's clearly valid about the present argument—especially since, in this case, the argument's validity *does* seem to have something to do with its form. For instance, the argument seems to have the same structure as the following valid argument:

P1 Froggy is a frog.
P2 All frogs are green.

C Froggy is green.

The first premise of the first argument says that some individual (Willard) is a member of a kind (logician). The second premise says that all members of that kind (logicians) have some property (they are pedantic). The conclusion says that the individual, therefore, also has that property (Willard is pedantic). Likewise, the first premise of the second argument says that an individual (Froggy) is a member of a kind (frogs). The second premise says that all members of that kind (frogs) have some property (greenness). The conclusion says that the individual, therefore, also has that property (Froggy is green).

What this tells us is that the method of symbolisation in $\mathcal{L}_S$ leaves out some important logical structure—it does not allow us to represent the **internal logical structure** of the atomic sentences which make up the premises and the conclusions of these two arguments. The *smallest unit of analysis* in $\mathcal{L}_S$ is the atomic sentence. Because of this, $\mathcal{L}_S$ is very good for symbolising the way in which atomic sentences can be joined together to form non-atomic sentences using truth-functional connectives. Sometimes, however, we need to consider the internal structure of the atomic sentences themselves in order to see what makes an argument have a valid form. It is therefore useful to have a way of representing the internal structure of atomic sentences. For this we need a more sophisticated formal language, which we'll label $\mathcal{L}_P$. The 'P' stands for *predicate*, which is a concept that we'll introduce shortly.

Luckily, while we're developing our new language $\mathcal{L}_P$, we can keep what we've learnt so far while developing $\mathcal{L}_S$. Coming up with a new formal language doesn't mean throwing out all our hard work with propositional logic and replacing it with something entirely different. Instead, we're going to *supplement* the old language with some new tools and techniques. In the end, we'll want a symbol system which includes *names*, *predicates*, *quantifiers* and *variables.*

1. Names and Individuals

For the purposes of logic, an **individual** is any specific person, place, or thing. So, for example, you and I are individuals, as are Frank Jackson and Mr. T. Each is an individual person, and hence an individual. London and Sydney are also individuals: they are individual cities, and (by virtue of this) they are individuals. This book is another kind of individual. And in the arguments above, Willard and Froggy were individuals.

In English, we have many different ways of referring to individuals. For example, we can use the *proper name* 'John Cleese' to refer to the individual John Cleese. Or, we could use a description which uniquely characterises John Cleese, such as 'the tallest member of Monty Python'. Descriptions which uniquely pick out an individual are called *definite descriptions*, and they usually have the form 'the so-and-so'. If he were around, we could refer to him by pointing and using a *demonstrative* like 'him' or 'that person there.' Or, if we were talking to him, we could refer to him using the simple pronoun 'you'.

In $\mathcal{L}_P$, we will use **names** to refer to individuals. We will use the italicised lower-case letters a through w for our names. Each must refer to exactly one individual. (We can add subscripts if more names needed.) In the jargon, these letters are also often called **constants** because they pick out just one individual; we'll refer to them as names and as constants.

We have set aside the symbols x, y, and z, which are *not* constants in $\mathcal{L}_P$. We reserve those letters for **variables**, which we will describe later when we introduce quantifiers.

> An INDIVIDUAL is a specific person, place, or thing. In $\mathcal{L}_P$, we use NAMES (also called CONSTANTS) to refer to specific individuals. Names are symbolised using italicised lower-case letters a through w.

A note: in English, we have lots of proper names that do not refer to any existing thing. For example—and, I'm sorry to break it to you—Santa Claus does not exist. So the name 'Santa Claus' is non-referring. Dealing with non-referring names is a tricky matter that philosophers debate. For our purposes, though, we don't have to worry about the problem of non-referring names. As we said above, every constant in our language $\mathcal{L}_P$ must refer to exactly one individual—no more, and *no less.* In other words, we will assume that there are no non-referring names in our logical language.

2. Predicates and Properties

Properties are attributes or qualities had by individuals. For instance, you and I have the property of *being human*, and neither of us have the property of *being green.* (I hope!) We also have the properties of *being living things*, *being mammals*, *being Earth-dwelling beings*, and so on. A dog shares many of these properties with us, but not all of them—clearly, a dog doesn't have the property of *being human*, and humans don't have the property of *being a dog.*

In English, we use **one-place predicates** refer to properties. So, for example, the phrase 'is red' is a one-place predicate that picks out the property of *being red.* (The reason for the 'one-place' will become clearer below, when we discuss relations.) In $\mathcal{L}_P$, we symbolise predicates using italicised capital letters; A, B, C, etc., with or without subscripts. Because of this, logics which use a formal language like $\mathcal{L}_P$ are sometimes called *predicate logic.* (We also used italicised capital letters in $\mathcal{L}_S$ to refer to whole sentences. However, there is no danger of confusion in also using these letters here for predicates, because in $\mathcal{L}_P$ we will not have any symbols for atomic sentences. Instead, every whole sentence will be symbolised in terms of its smaller sub-sentential parts.)

So we now have symbols that represent properties and individuals. We can use these to symbolise some simple sentences.

1. Jane is angry.
2. Bob is red.
3. Bob is angry.
4. Jane is angry and Bob is red.
5. If Jane is angry then Jane is red.

Let's use the following symbolisation key:

j : Jane
b : Bob
A : ... is angry
R : ... is red

By convention, we always put the predicate out the front of the sentence, and we will use the same connectives as we used in previous chapters. Hence:

Jane is angry $\Longrightarrow Aj$
Bob is red $\Longrightarrow Rb$
Bob is angry $\Longrightarrow Ab$
Jane is angry and Bob is red $\Longrightarrow Aj \wedge Rb$
If Jane is angry then Jane is red $\Longrightarrow Aj \rightarrow Rj$

3. Relations

Besides properties, we also need a way of talking about **relations**. Relations are ways in which individuals can be connected to one another. A **two-place** (or **dyadic**, or **binary**) relation connects two individuals. For example, if we say that 'Frank is married to Jane', we are saying that Frank is connected to Jane via the *is married to* relation. And if we say 'Jane is to the left of Frank', we are saying that Jane is connected to Frank via the *is to the left of* relation. Note that individuals can also be related to themselves in various ways. For example, 'Frank is identical to Frank' just means that Frank is connected to himself via the *identity* relation.

We can also have **three-place** (or **triadic**, or **ternary**) relations. For example, to say that 'Frank sits between Bob and Jane' is to say that Frank is related to Bob and Jane via the *sits between* relation. Similarly, '2 is the sum of 1 and 1'; in this case, we're saying that the number 2 is related to 1 and 1—the same individual, twice over—via the *is the sum of* relation.

For any natural number n, we can have ***n*-place** (or ***n*-adic**, or ***n*-ary**) relations—though natural examples of n-adic relations for large values of n become progressively more difficult to find. A four-place relation which sometimes gets used in mathematics is *the difference between ... and ... is greater than the difference between ... and ...*; there are not many examples of 5-adic relations.

In English, ***n*-place predicates** are used to refer to n-place relations. As such, in $\mathcal{L}_P$, we use the same kind of symbols to refer to both (one-place) properties and (n-place) relations. The difference is that an n-place predicate will be followed by n lower-case letters. This is sometimes described by saying that an n-place predicate has n **arguments**—an "argument" in this context can be thought of as a 'slot' for the placement of a name.

> PROPERTIES are attributes or qualities had by individuals, and RELATIONS are ways in which individuals can be connected to one another. An n-place relation connects n-many (not necessarily distinct) individuals. In $\mathcal{L}_P$, we use PREDICATES to refer both to properties and to relations.

We can now symbolise the following sentences in $\mathcal{L}_P$:

6. Abbie is to the left of Brett.
7. Abbie is identical to Abbie.
8. Abbie is identical to Caroline.
9. Abbie is between Brett and Caroline.
10. Abbie is to the left of Brett, and between Brett and Caroline.

We use this key:

a : Abbie
b : Brett
c : Caroline
L : ... is to the left of ...
I : ... is identical to ...
B : ... is between ... and ...

We then symbolise Sentence 6 as Lab. Notice that the order of the constants matters: because we put the a first, Lab says that Abbie is to the left of Bob. If we had instead put down Lba, we'd have that Bob is to the left of Abbie. In the case of Sentence 7, we should have the symbolisation Iaa. Here, the order of the constants does not matter—but only because we have used the same name twice.

In the case of Sentence 8, we should translate it as Iac. In this case, we're using two distinct names (a and c) to refer to the same individual. This is the same as it is in ordinary English: every proper name refers to one and only one individual, but a single individual can have many different names.

Betweenness is a three-place relation, so we can symbolise Sentence 9 as $Babc$. In this case, Abbie (a) is the one who is B-related to Bob (b) and Caroline (c), so the a should be the constant placed furthest to the left.

Finally, we can symbolise Sentence 10 as $Lab \land Babc$. Even though the name 'Abbie' does not appear after the word 'and', the meaning of Sentence 10 is the same as '[Abbie is to the left of Bob] *and* [Abbie is between Bob and Caroline].'

4. Quantifiers and Variables

We are now ready to introduce quantifiers. **Quantifiers** are words like 'every' and 'some'. There are many quantifiers besides these (e.g., 'most'), but in $\mathcal{L}_P$ we only symbolise 'every' and 'some'. Later on in the book, we will see how we can define up many other quantifiers in terms of 'every' and 'some. Indeed, we'll see how we can define 'every' in terms of 'some', and 'some' in terms of 'every'.

Consider these sentences:

11. Everything is red.
12. Everything is to the left of Bob.

With the following symbolisation key:

b : Bob
R : ... is red
L : ... is to the left of ...

How could we translate Sentence 11? If we had at least one name for every single individual of whom we might possibly be speaking, we could try to translate it as $Ra \land (Rb \land (Rc \land (Rd \dots)))$. This isn't very convenient. Indeed, if there were infinitely many people, we'd need infinitely many names and infinitely many conjuncts.

It would be nice to have a simpler way of symbolising Sentence 11. In order to do this, we introduce the '$\forall$' symbol. This is called the **universal quantifier**. There are a number of slightly different notational conventions for dealing with quantifiers, but we will use the following: a quantifier must always be followed by a single **variable**, and then a formula which includes that variable (and which must be enclosed in brackets). A **formula** is a string of symbols which can be used to represent a proper sentence once all of the variables within it have been replaced by constants.

An example will help. We symbolise Sentence 11 as $\forall x(Rx)$. We can read this as saying 'For all things x: x is red.' The universal quantifier $\forall$ is followed immediately by the variable x, and then a formula Rx enclosed in brackets. Rx is called a formula because if we were to replace the variable x with a constant (like b), the result (Rb) would symbolise a complete sentence—in this case, 'Bob is red'. The bracketing around the formula Rx delineates the *scope* of the quantifier that comes before it. We will give a formal definition of scope later, but intuitively it is the part of the sentence that the quantifier applies to.

Sentence 14 can be paraphrased as, 'For all things x: x is to the left of Bob.' Symbolised it becomes $\forall x(Lxb)$. In this case, the $\forall$ quantifier takes scope over the formula Lxb. Just as with Sentence 11, the variable x acts as a "stand in" for the (possibility infinite) constants that could name all the many different individuals.

Note that there is no special reason to use the variable x to symbolise Sentence 11 or Sentence 14, rather than the other variables y or z. For example $\forall x(Rx)$ means exactly the same thing as $\forall y(Ry)$, which means the same thing as $\forall z(Rz)$. Usually, when more than one variable is needed in a given symbolisation, it's easiest to use x for the first variable, y for the second, and z for the third. But this is just for bookkeeping purposes.

> The symbol $\forall$ is the UNIVERSAL QUANTIFIER. A universal quantifier is always followed by a variable (x, y, or z), and then a formula which includes that variable, and which is enclosed in brackets. If $\mathcal{F}$ is any formula, then $\forall x(\mathcal{F})$ can be read as saying 'For all things x: $\mathcal{F}$.' For example, if $\mathcal{F}$ is the formula Px, then $\forall x(Px)$ can be read as 'For all things x: x is P.

Two new sentences to consider:

13. Something is red.
14. Bob is to the left of something.

To symbolise Sentence 13, we can introduce another new symbol: the **existential quantifier**, $\exists$. Like the universal quantifier, the existential quantifier must always be followed by a variable and then a formula which includes that variable and is enclosed in brackets. Thus, Sentence 13 can be translated as $\exists x(Rx)$. This can be read as 'There exists at least one x such that: x is red.' And Sentence 14 can be symbolised as $\exists x(Lbx)$.

> The symbol $\exists$ represents the EXISTENTIAL QUANTIFIER. An existential quantifier is always followed by a variable (x, y, or z), and then a formula which includes that variable, and which is enclosed in brackets. If $\mathcal{F}$ is any formula, then '$\exists x(\mathcal{F})$' can be read as saying 'There exists at least one thing x such that: $\mathcal{F}$'. For example, if $\mathcal{F}$ is the formula Px, then $\exists x(Px)$ can be read as 'There exists at least one thing x such that: x is P.

Now let's consider these further sentences:

15. Nothing is red.
16. There is something which is not red.
17. Not everything is red.

Sentence 15 can be paraphrased as, 'It is not the case that something is red.' In other words, it is the negation of Sentence 13, and can therefore be translated as $\neg\exists x(Rx)$. Sentence 15 can *also* be paraphrased as, 'Everything is not red.' In that case, we might want to symbolise it as $\forall x(\neg Rx)$.

Is this a problem? Not at all! In fact, both are perfectly acceptable symbolisations of the same sentence, because they are logically equivalent. In general, the following logical equivalences always hold:

$\forall x(\mathcal{F})$ if and only if $\neg\exists x(\neg\mathcal{F})$
$\exists x(\mathcal{F})$ if and only if $\neg\forall x(\neg\mathcal{F})$
$\forall x(\neg\mathcal{F})$ if and only if $\neg\exists x(\mathcal{F})$
$\exists x(\neg\mathcal{F})$ if and only if $\neg\forall x(\mathcal{F})$

This means that any declarative sentence in English which can be symbolised with a universal quantifier can also be symbolised with an existential quantifier, and *vice versa*. One particular means of symbolising might seem more natural than the other on a given occasion, but there is no logical difference in translating with one quantifier rather than the other. For some sentences, it will simply be a matter of taste. Strictly speaking, our formal language really only needs only one of either $\forall$ or $\exists$. However, we make things somewhat easier for ourselves if we have two quantifiers, even if one of them is redundant given the existence of the other.

Sentence 16 is most naturally paraphrased as, 'There is at least one thing x such that: x is not red.' This becomes $\exists x(\neg Rx)$. Equivalently, we could write it as $\neg\forall x(Rx)$; i.e., 'It's not the case that [everything is red].' That's just what Sentence 17 also says, so Sentence 16 is logically equivalent to Sentence 17

5. The Universe of Discourse

Given the symbolisation key we have been using, '$\forall x(Rx)$' means 'Everything is red.' What if we wanted to say that 'Everyone is red,' where 'everyone' refers only to the people in the room? There are two main options here.

The first option is to introduce a symbol for the predicate "... is a person in this room" and paraphrase 'Everyone is red' as 'Every person in this room is red'.

P : ... is a person in this room
R : ... is red

With this symbolisation key, we can translate 'Every person in this room is red' as $\forall x(Px \rightarrow Rx)$. This says 'For all things x: if x is a person in this room, then x is red.' In this case, the antecedent tells us that of all the individuals that exist, we are only interested in people currently in the room; the consequent then says that each of *those* individuals are red.

In general, we can translate any sentence which means something of the form 'Every thing which is P is Q' as $\forall x(Px \to Qx)$. Similarly, we can translate any sentence which means something of the form 'All things which are P are Q' as $\forall x(Px \to Qx)$. (In English we recognise an 'is'/'are' distinction which is not as important when we translate into $\mathcal{L}_P$.)

In this case, the $\forall x$ refers to *everything*; i.e., every individual that exists, anywhere. That is a huge number of individuals—indeed, infinitely many. If we wanted to, however, we could stipulate that $\forall x$ refers to a more restricted class of individuals. For example, we could just make it mean "every person in this room." In that case, when we write $\forall x(Rx)$, we would be saying 'For every person x in this room: x is red.' To alter the meaning of $\forall x$ in this way, we need to specify what we'll call the **universe of discourse**, abbreviated UoD. The UoD contains every individual that we might potentially be talking about *in the present context.* It's up to us to specify which UoD we intend, and it can be almost anything we like. The only restriction in $\mathcal{L}_P$ is that the UoD must be *non-empty*; that is, it must contain at least one individual. But besides that rule, we're free to choose our own universe of discourse.

The universe of discourse could be taken to be everything *simpliciter.* In that case, we would need to translate 'Everyone is red' as $\forall x(Px \to Rx)$, as above. There are many things this UoD which aren't people, so it would be wrong in this case to symbolise the sentence as $\forall x(Rx)$. However, we could also stipulate that, for present purposes, the UoD is just the set of all people in the room. In this case, $\forall x(Rx)$ is perfectly appropriate as a translation of 'Everyone is red.' Logicians would say that in each case, the variable x *ranges over* a restricted universe of discourse.

Strictly speaking, then, $\forall x(\mathcal{F})$ should really be interpreted as 'For every individual x *in the universe of discourse*: $\mathcal{F}$)'. Similarly, $\exists x(\mathcal{F})$ should be read as 'There is at least one individual x *in the universe of discourse* such that: $\mathcal{F}$.' Usually, the intended restrictions on the universe of discourse are left implicit, and can easily be discovered from context. For example, when we use words like 'everybody', 'everyone', 'anyone', 'somebody', 'someone', and 'no one', we are often restricting the universe of discourse to some salient set of people. When we say 'everywhere', 'somewhere', and 'nowhere', we are restricting the universe of discourse to a set of locations. And, quite often when we say 'Everything' and 'nothing', we don't mean *everything in the universe*, just *everything of a certain kind around here.*

Restricted quantification occurs frequently in natural language, and we are very good at interpreting it when it happens. When we're doing logic, it's often helpful to avoid the possibility of ambiguity. So, whenever there is some potential for confusion, you should remember to specify the intended UoD.

6. Symbolisation

We now have all of the pieces needed to characterise the logical language $\mathcal{L}_P$. Translating more complicated sentences will only be a matter of knowing the right way to combine predicates, constants, quantifiers, variables, and connectives. In this section, we'll look at sentences which only make use of a single quantifier. Sentences with multiple quantifiers are usually much harder to deal with, and are covered in the next chapter. Consider:

18. Every coin on the table is a dime.
19. Some coin on the table is a dime.
20. Not all the coins on the table are dimes.
21. None of the coins in my pocket are dimes.

In providing a symbolisation key, we will henceforth need to specify a UoD. Since we are talking about coins in my pocket and on the table, the UoD must *at least* contain all of those coins. Since we are not talking about anything besides coins, we let the UoD be all coins. Since we are not talking about any specific coins, we do not need to define any constants. So we define this key:

UoD : All coins
P : ... is in my pocket.
T : ... is on the table.
Q : ... is a quarter.
D : ... is a dime.

Sentence 18 is most naturally translated with a universal quantifier. The universal quantifier says something about everything in the UoD, not just about the coins on the table. Sentence 18 means that (for any coin), *if* that coin is on the table, *then* it is a dime. So we can translate it as $\forall x(Tx \rightarrow Dx)$. Sentence 19 is most naturally translated with an existential quantifier. It says that there is some coin which is both on the table and which is a dime. So we can translate it as $\exists x(Tx \wedge Dx)$.

(Notice that we needed to use a conditional with the universal quantifier in Sentence 18, but we used a conjunction with the existential quantifier in Sentence 19. What would it mean to write $\exists x(Tx \rightarrow Dx)$? Probably not what you think. It means that there is some individual (coin) in the UoD which would satisfy the formula $(Tx \rightarrow Dx)$—there is some coin a such that $(Ta \rightarrow Da)$ is true. In our earlier language $\mathcal{L}_S$, $(\mathcal{X} \rightarrow \mathcal{Y})$ is logically equivalent to $(\neg\mathcal{X} \vee \mathcal{Y})$, and this also holds in $\mathcal{L}_P$. So $\exists x(Tx \rightarrow Dx)$ is true if there is some a such that $(\neg Ta \vee Da)$ holds—that is, it is true if some coin is *either* not on the table *or* is a dime. A conditional will usually be the natural connective to use with a universal quantifier, but a conditional within the scope of an existential quantifier can do very strange things. As a general rule, do not put conditionals in the scope of existential quantifiers unless you are certain that you need one to say what you want.)

Sentence 20 can be paraphrased as, 'It is not the case that every coin on the table is a dime.' So we can translate it as $\neg\forall x(Tx \rightarrow Dx)$. You might look at Sentence 20 and paraphrase it instead as, 'Some coin on the table is not a dime.' You would then symbolise it as $\exists x(Tx \wedge \neg Dx)$. Although it is probably not obvious, these two translations are logically equivalent. This is due to the logical equivalence between $\neg\forall x(\mathcal{F})$ and $\exists x(\neg\mathcal{F})$, along with the equivalence between $\neg(\mathcal{X} \rightarrow \mathcal{Y})$ and $(\mathcal{X} \wedge \neg\mathcal{Y})$.

Sentence 21 can be paraphrased as, 'It is not the case that there is some dime in my pocket.' This can be symbolised as $\neg\exists x(Px \wedge Dx)$. It might also be paraphrased as, 'Everything in my pocket is not a dime,' and then could be symbolised as $\forall x(Px \rightarrow \neg Dx)$. Again the two translations are logically equivalent. Both are perfectly adequate ways of symbolising Sentence 21, and you should feel free to choose whichever feels more natural to you.

We can now translate the arguments from p. 54, the ones that motivated the need for $\mathcal{L}_P$ in the first place:

P1 Willard is a logician.
P2 All logicians are pedantic.

C Willard is pedantic.

P1 Froggy is a Frog.
P2 All frogs are green.

C Froggy is green.

We'll assume that the UoD for both arguments is *everything*, and use the following symbolisation keys:

L :	... is a logician	F :	... is a frog
P :	... is pedantic	G :	... is green
w :	Willard	f :	Froggy

P1 Lw
P2 $\forall x(Lx \rightarrow Px)$

C Pw

P1 Ff
P2 $\forall x(Fx \rightarrow Gx)$

C Gf

As you can see, we've managed to capture the structure that the two arguments share, in a way that wasn't possible to do using $\mathcal{L}_S$. The only thing that differs between the arguments is the particular predicates and names involved, but the basic *internal structure* of their premises and conclusions are the same.

Chapter 5: Key Ideas

- ▷ When symbolising in $\mathcal{L}_P$, we consider not only how sentences are constructed out of other sentences using connectives, but also the internal structure of atomic sentences. Sentences in $\mathcal{L}_P$ are constructed out of **names** (a, b, c, …, w), **predicates** (A, B, C, …), **quantifiers** ($\forall$, $\exists$), **variables** (x, y, z), **connectives** ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$), and brackets.

- ▷ Names (also known as **constants**) refer to individual people, places, or things. A predicate can be **1-place**, or **2-place**, or ***n*-place** for any n. A 1-place predicate refers to a property, like *… is red*; a 2-place predicate refers to a 2-place relation, like *… is taller than …*; a 3-place predicate refers to a 3-place relation, and so on.

- ▷ By convention, the predicate always comes first. So if the names b and c refer to the individual Bob and Caroline respective, and the predicate H refers to the property *… is happy*, then we symbolise 'Bob is happy' as Hb and 'Caroline is not happy' as $\neg Hc$. The order of the names is important. If T refers to the relation *… is taller than…*, then we symbolise 'Bob is taller than Caroline' as Hbc, whereas we symbolise 'Caroline is not taller than Bob' as $\neg Hcb$.

- ▷ The symbol $\forall$ is the **universal quantifier**. A universal quantifier is always followed by a variable (x, y, or z), and then a formula which includes that variable, and which is enclosed in brackets. If $\mathcal{F}$ is any formula, then $\forall x(\mathcal{F})$ can be read as saying 'For all things x: $\mathcal{F}$.' For example, if $\mathcal{F}$ is the formula Px, then $\forall x(Px)$ can be read as 'For all things x: x is P.

- ▷ The symbol $\exists$ represents the **existential quantifier**. An existential quantifier is always followed by a variable (x, y, or z), and then a formula which includes that variable, and which is enclosed in brackets. If $\mathcal{F}$ is any formula, then '$\exists x(\mathcal{F})$' can be read as saying 'There exists at least one thing x such that: $\mathcal{F}$'. For example, if $\mathcal{F}$ is the formula Px, then $\exists x(Px)$ can be read as 'There exists at least one thing x such that: x is P.

- ▷ $\forall x(\mathcal{F})$ is always logically equivalent to $\neg\exists x(\neg\mathcal{F})$, and $\exists x(\mathcal{F})$ is always logically equivalent to $\neg\forall x(\neg\mathcal{F})$. In general, any sentence with a universal quantifier can be converted into a logically equivalent sentence with an existential quantifier, and vice versa.

- ▷ The **Universe of Discourse** (UoD) specifies the set of individuals over which $\forall$ and $\exists$ quantify. For example, if the Universe of Discourse is stipulated to be "All human beings", then $\forall x(Px)$ means 'For all human beings x: x is P.'

Practice Exercises

$\star$ **Part A** Using the symbolisation key given, symbolise each of the following into its nearest $\mathcal{L}_P$ equivalent.

$$\begin{aligned}
\text{UoD}&: \text{ The set of all animals}\\
A&: \text{ \ldots is an alligator}\\
M&: \text{ \ldots is a monkey}\\
R&: \text{ \ldots is a reptile}\\
Z&: \text{ \ldots lives at the zoo}\\
L&: \text{ \ldots loves \ldots}\\
a&: \text{ Amos}\\
b&: \text{ Bouncer}\\
c&: \text{ Cleo}
\end{aligned}$$

1. Amos, Bouncer, and Cleo all live at the zoo.
2. Bouncer is a reptile, but not an alligator.
3. If Cleo loves Bouncer, then Bouncer is a monkey.
4. If both Bouncer and Cleo are alligators, then Amos loves them both.
5. Some reptile lives at the zoo.
6. Every alligator is a reptile.
7. There are reptiles which are not alligators.
8. Cleo loves a reptile.
9. Bouncer loves all the monkeys that live at the zoo.

$\star$ **Part B** Using the symbolisation key given, symbolise each of the following into its nearest $\mathcal{L}_P$ equivalent.

$$\begin{aligned}
\text{UoD}&: \text{ The set of all candies}\\
C&: \text{ \ldots has chocolate in it}\\
M&: \text{ \ldots has marzipan in it}\\
S&: \text{ \ldots has sugar in it}\\
T&: \text{ Boris has tried \ldots}\\
B&: \text{ \ldots is better than \ldots}
\end{aligned}$$

1. Boris has never tried any candy.
2. Marzipan is always made with sugar.
3. Some candy is sugar-free.
4. Every sugar-free candy has marzipan in it.
5. No candy is better than itself.
6. Boris has never tried sugar-free chocolate.

Part C
Each of the answers for **Part B** can be expressed using either an existential quantifier or a universal quantifier. So, for this exercise, for each of the sentences in **Part B**, (a) if you used an existential quantifier then try answering it again using a universal quantifier, and (b) if you used a universal quantifier then try answering it again using an existential quantifier.

Part D These are syllogistic figures identified by Aristotle and his successors, along with their medieval names. Translate each argument into $\mathcal{L}_P$.

Barbara: All Bs are Cs. All As are Bs. Therefore, all As are Cs.
Baroco: All Cs are Bs. Some A is not B. Therefore, some A is not C.
Bocardo: Some B is not C. All As are Bs. Therefore, some A is not C.
Celantes: No Bs are Cs. All As are Bs. Therefore, no Cs are As.
Celarent: No Bs are Cs. All As are Bs. Therefore, no As are Cs.
Cemestres: No Cs are Bs. No As are Bs. Therefore, no As are Cs.
Cesare: No Cs are Bs. All As are Bs. Therefore, no As are Cs.
Dabitis: All Bs are Cs. Some A is B. Therefore, some C is A.
Darii: All Bs are Cs. Some A is B. Therefore, some A is C.
Datisi: All Bs are Cs. Some A is B. Therefore, some A is C.
Disamis: Some B is C. All As are Bs. Therefore, some A is C.
Ferison: No Bs are Cs. Some A is B. Therefore, some A is not C.
Ferio: No Bs are Cs. Some A is B. Therefore, some A is not C.
Festino: No Cs are Bs. Some A is B. Therefore, some A is not C.
Baralipton: All Bs are Cs. All As are Bs. Therefore, some C is A.
Frisesomorum: Some B is C. No As are Bs. Therefore, some C is not A.

Part E Using the symbolisation key given, symbolise each of the following into its nearest $\mathcal{L}_P$ equivalent.

UoD :	The set of all animals
D :	. . . is a dog
S :	. . . likes samurai movies
L :	. . . is larger than . . .
b :	Bertie
e :	Emerson
f :	Fergis

1. Bertie is a dog who likes samurai movies.
2. Bertie, Emerson, and Fergis are all dogs.
3. Emerson is larger than Bertie, and Fergis is larger than Emerson.
4. All dogs like samurai movies.
5. Only dogs like samurai movies.
6. There is a dog that is larger than Emerson.
7. If there is a dog larger than Fergis, then there is a dog larger than Emerson.
8. No animal that likes samurai movies is larger than Emerson.
9. No dog is larger than Fergis.
10. Any animal that dislikes samurai movies is larger than Bertie.
11. There is an animal that is between Bertie and Emerson in size.
12. There is no dog that is between Bertie and Emerson in size.
13. No dog is larger than itself.

Chapter 6

Advanced Symbolisation

In this chapter, we will look at a few tricky issues that arise when we're translating English language sentences into $\mathcal{L}_P$. The issues covered in this chapter are a bit more advanced than those that come before.

1. Empty Predicates and a fallacy

Nothing we said in the previous chapter guarantees that a given predicate will always apply to something in the Universe of Discourse. A predicate that applies to nothing in the UoD is called an **empty predicate**. For example, suppose we want to symbolise these two sentences:

1. Every monkey knows sign language.
2. Some monkey knows sign language.

It is possible to write the symbolisation key for these sentences in this way:

M : ... is a monkey
S : ... knows sign language

Sentence 1 can now be translated as $\forall x(Mx \to Sx)$, and Sentence 2 becomes $\exists x(Mx \wedge Sx)$. It is tempting to think that Sentence 1 entails Sentence 2. That is: if every monkey knows sign language, then it must be the case that some monkey knows sign language. Indeed, Aristotle thought that the argument had a valid form, which in his system was characterised along the following lines:

P1 All Ms are S.

C Some M is S.

However, the entailment does *not* hold in $\mathcal{L}_P$. It is possible for the sentence $\forall x(Mx \to Sx)$ to be true even though the sentence $\exists x(Mx \wedge Sx)$ is false.

How can this be? The answer comes from considering whether the premise and the conclusion would be true *if there were no monkeys.* We have characterised $\forall$ and $\exists$ in such a way that $\forall x(\mathcal{F})$ is equivalent to $\neg\exists x(\neg\mathcal{F})$. As such, the universal quantifier doesn't imply the existence of anything—it only implies the non-existence of certain things. If Sentence

1 is true, then there are *no* monkeys who don't know sign language. Now, if there were no monkeys, then $\forall x(Mx \rightarrow Sx)$ would be true. Hence, $\exists x(Mx \wedge Sx)$ would be false. So Sentence 1 does not imply Sentence 2 after all. If there was at least one monkey, the entailment would go through—but logic alone does not tell us whether any monkeys exist.

It is important to allow empty predicates into our language, because we want to be able to say things like, 'I do not know if there are any monkeys, but any monkeys that there are must know sign language.' That is, we want to be able to speak about properties that do not (or might not) belong to anything. Alternatively, consider the property of *being a unicorn*. We might use U to refer to this property. We can then translate 'There are no unicorns' as $\neg\exists x(Ux)$; i.e., 'It is not the case that there is at least one x such that: x is a unicorn'. We would find this very difficult to say if we didn't allow empty predicates into our formal language. Indeed, our ability to formulate such sentences is a huge advantage of $\mathcal{L}_P$; philosophers of the past worried a great deal about how we can make sense of sentences like 'There are no unicorns.'

Once we have empty predicates in our language, some interesting things happen. For example, consider the sentence 'Every unicorn knows sign language.' Translated, this becomes $\forall x(Ux \rightarrow Sx)$, which is *true*. This is counterintuitive, since we do not want to say that there are a whole bunch of unicorns out there that know sign language. It is important to remember, though, that $\forall x(Ux \rightarrow Sx)$ just means that any member of the UoD which is a unicorn is also something that knows sign language. Since there are no unicorns in the UoD, the sentence is trivially true.

2. Translating pronouns

When translating into $\mathcal{L}_P$, it is important to understand the structure of the sentences you want to translate. What matters is the final translation in $\mathcal{L}_P$, and sometimes you will be able to move from an English language sentence directly to a sentence of $\mathcal{L}_P$. Other times, it helps to paraphrase the sentence one or more times. Each successive paraphrase should move from the original sentence closer to something that you can translate directly into $\mathcal{L}_P$.

For the next several examples, we will use this symbolisation key:

UoD : The set of all people
G : ... can play guitar
R : ... is a rock star
l : Lemmy

Now consider these sentences:

3. If Lemmy can play guitar, then he is a rock star.
4. If a person can play guitar, then he is a rock star.

Sentence 3 and Sentence 4 have the same consequent ('he is a rock star'), but they cannot be translated in the same way. It helps to paraphrase the original sentences, replacing pronouns with explicit references. Sentence 3 can be paraphrased as, 'If Lemmy can play guitar, then *Lemmy* is a rockstar.' This can be translated as $(Gl \rightarrow Rl)$. We do not need any quantifiers for this. On the other hand, Sentence 4 must be paraphrased differently: 'If a person can play guitar, then *that person* is a rock star.' This sentence is not about any particular person, so we need a variable. We can re-paraphrase the sentence as, 'For any person x: if x can play guitar, then x is a rockstar.' Now this can be translated as $\forall x(Gx \rightarrow Rx)$. This says just the same thing as 'Every person who can play guitar is a rock star.'

Consider now these other sentences:

5. If anyone can play guitar, then Lemmy can.
6. If anyone can play guitar, then he or she is a rock star.

These two sentences have the same antecedent ('If anyone can play guitar'), but they have different logical structures.

Sentence 5 can be paraphrased, 'If someone can play guitar, then Lemmy can play guitar.' The antecedent and consequent are separate sentences, so it can be symbolised with a conditional as the main logical operator: $\exists x(Gx) \rightarrow Gl$. That is, 'If there exists an x such that: x can play guitar, then Lemmy can play the guitar. On the other hand, Sentence 6 can be paraphrased, 'For any particular person, if that person can play guitar, then he or she is a rock star.' It would be a mistake to symbolise this with an existential quantifier, because it is talking about everybody. The sentence is equivalent to 'All guitar players are rock stars.' It is best translated as $\forall x(Gx \rightarrow Rx)$.

The English words 'any' and 'anyone' should typically be translated using quantifiers. As these two examples show, they sometimes call for an existential quantifier (as in Sentence 5) and sometimes for a universal quantifier (as in Sentence 6). If you have a hard time determining which is required, paraphrase the sentence with an English language sentence that uses words besides 'any' or 'anyone.'

3. Ambiguous predicates

Suppose we want to symbolise this argument:

P1 Adina is a doctor, but is not skilled.
P2 All doctors play tennis.
P3 The hospital will hire a skilled doctor.

C Adina plays tennis, but the hospital will not hire her.

Given the structure of the first premise, you might think that we should use the following symbolisation key:

UoD :	The set of all people
D :	. . . is a doctor
S :	. . . is skilled
T :	. . . is a tennis player
H :	The hospital will hire . . .
a :	Adina

In that case, you might symbolise the argument like so:

P1 $Da \land \neg Sa$
P2 $\forall x(Dx \to Tx)$
P3 $\forall x(Hx \to (Dx \land Sx))$

C $Tb \land \neg Hb$

But this is not quite right. Being a skilled doctor is not the same thing as being skilled and being a doctor. For example, a doctor who is very skilled at tennis is skilled, but they are not therefore a skilled doctor.

Another way to see that something has gone wrong with the previous translation is to consider what happens when we apply the same method to the following argument:

P1 Adina is a skilled doctor and a tennis player.

C Therefore, Adina is a skilled tennis player.

This argument is clearly invalid. However, look what happens when we translate 'Adina is a skilled doctor' into $Da \land Sa$, and 'Adina is a skilled tennis player' into $Ta \land Sa$, as we did when translating the former argument:

P1 $(Da \land Sa) \land Ta$

C $Sa \land Ta$

This argument *is* formally valid, so there must have been some mistake in the translation. The problem is that there is a difference between being *skilled as a doctor* and *skilled as a tennis player*. Translating the previous arguments correctly requires us to introduce separate predicates, one for each type of skill The following key uses X for being a skilled doctor, and Y for being a skilled tennis player:

UoD : The set of all people
D : ... is a doctor
X : ... is skilled doctor
Y : ... is skilled tennis player
T : ... is a tennis player
H : The hospital will hire ...
a : Adina

Using this as our key, the first argument becomes:

P1 $Da \land \neg Xa$
P2 $\forall x(Dx \to Tx)$
P3 $\forall x(Hx \to Xx)$

C $Tb \land \neg Hb$

And the second argument becomes:

P1 $Xa \wedge Ta$

C Ya

As you can see, this argument is formally invalid, so the translation this time looks correct.

The moral of these examples is that you need to be careful when symbolising predicates, and ensure you don't do so in an ambiguous way. Similar problems can arise with predicates like *skilled*, which might mean many different things in different contexts. The same goes for *good*, *bad*, *big*, and *small*, and many other examples. Just as skilled doctors and skilled tennis players have different skills, big dogs, big mice, and big problems are big in different ways.

4. Multiple Quantifiers

Consider this following symbolisation key and the sentences that follow it:

F: ... is a friend of ...
U: ... is unhappy with ...
g: Gerald

7. Gerald is the friend of someone.
8. Someone is Gerald's friend.
9. Someone has a friend.
10. All of Gerald's friends are unhappy with everyone.
11. Someone is friends with everyone.
12. Everyone is friends with someone.

Sentence 7 can be paraphrased as 'There is an x such that: Gerald is x's friend. Thus, we can translate it as $\exists x(Fgx)$. This is not to be confused with the correct translation of Sentence 8, which is $\exists x(Fxg)$.

Sentence 9 can be paraphrased as 'There exists an x such that: x is the friend of someone. Note that the subsentence, 'x is a friend of someone,' is just the same as Sentence 7 but with 'Gerald' replaced by the variable x. Alternatively, Sentence 9 could also be paraphrased as 'There exists a y such that: someone is y's friend. In this case, the subsentence 'someone is y's friend' is just the same as Sentence 8, but with 'Gerald' replaced by the variable y.

So on both ways of paraphrasing the sentence, the sub-sentences that follow the quantifier contain yet another quantifier. In order to symbolise Sentence 9, then, we will need to make use of multiple quantifiers. Thus, we can write it as $\exists x\exists y(Fxy)$. This can be read as 'There exists an x and there exists a y such that: x is the friend of y.'

Sentence 10 can be paraphrased as 'For all x: if x is a friend of Gerald, then x is unhappy with everyone. It could be equally well paraphrased as 'For all y: all of Gerald's friends are unhappy with y. So, again, we're going to need to make use of multiple quantifiers. We can write it as $\forall x\forall y(Fxg \rightarrow Uxy)$. In semi-formal English, this can be read as 'For all x and all y: *if* x is friends with Gerald, *then* x is unhappy with y.'

What about Sentence 11? Here, we're going to need to start mixing quantifiers. We can paraphrase Sentence 11 as 'There is an x such that: x is friends with everyone.' The subsentence 'x is friends with everyone' involves a universal quantifier, so we can alternatively paraphrase Sentence 11 as 'There is an x such that: for all y: x is friends with y.' Thus, we can symbolise Sentence 11 as $\exists x\forall y(Fxy)$.

Note that the order in which the quantifiers appear in our symbolisation of Sentence 11. When we're mixing quantifiers, the order matters. To see this, compare Sentence 11 with Sentence 12. Here, the paraphrase would be: 'For all x: there exists a y such that: x is friends with y.' In other words, we should translate Sentence 12 as $\forall x\exists y(Fxy)$.

Let's now consider this symbolisation key and the sentences that follow it:

UoD :	The set of all people
L :	. . . likes . . .
i :	Imre
k :	Karl

13. Imre likes everyone that Karl likes.
14. There is someone who likes everyone who likes everyone that he likes.

Sentence 13 can be partially translated as $\forall x$(Karl likes $x \rightarrow$ Imre likes x). This becomes $\forall x(Lkx \rightarrow Lix)$. On the other hand, Sentence 14 is almost a tongue-twister. There is little hope of writing down the whole translation immediately, but we can proceed by small steps. An initial, partial translation might look like this:

$$\exists x \text{ (everyone who likes everyone that } x \text{ likes is liked by } x)$$

The part that remains in English is a universal sentence, so we translate further:

$$\exists x\forall y(\text{if } y \text{ likes everyone that } x \text{ likes, then } x \text{ likes } y)$$

Now the consequent of the conditional is straightforward to symbolise: Lxy. Notice also that the antecedent is structurally just like Sentence 13, but with y and x in place of i and k. So, Sentence 14 can be completely translated in this way

$$\exists x\forall y(\forall z(Lxz \rightarrow Lyz) \rightarrow Lxy)$$

When symbolising sentences with multiple quantifiers, it is best to proceed by small steps. In the first step, paraphrase until you come up with an English sentence so that the logical structure is readily symbolised in $\mathcal{L}_P$. Then, translate piecemeal, replacing the daunting task of translating a long sentence with the simpler task of translating shorter formulae.

5. Identity

Consider this sentence:

15. Pavel owes money to everyone else.

Let the UoD be the set of all people; this will allow us to translate 'everyone' as a universal quantifier. Let O mean . . . owes money to . . ., and let p refer to Pavel. Now we can symbolise Sentence 15 as $\forall x(Opx)$. Unfortunately, this translation has some odd consequences. It says that Pavel owes money to every member of the UoD, including Pavel himself. However, Sentence 15 does not say that Pavel owes money to himself; he owes money to everyone *else*. This is a problem, because $\forall x(Opx)$ is the best translation we can give of this sentence into $\mathcal{L}_P$.

It can be exceedingly useful to treat identity as a distinctive logical symbol all of its own, rather than treating it like any other 2-place predicate. Since it has a special logical meaning, we use $=$ a bit differently than we've been using other predicates: instead of placing its two arguments after it, we will place the arguments on either side of it. So, we should read $a = b$ as meaning a is identical to b. This does not mean merely that a and b are indistinguishable or that all of the same predicates are true of them. Rather, it means that a and b are *the very same thing*.

When we write $a \neq b$, we mean that a and b are not identical.' There is no reason to introduce this as an additional predicate. Instead, $a \neq b$ is simply an abbreviation of $\neg(a = b)$. Now suppose we want to symbolise this sentence:

16. Pavel is Mister Checkov.

Let the constant c mean Mister Checkov. Then, 16 can be symbolised as $p = c$. This means that the constants p and c both refer to the same person. This is all well and good, but how does it help with Sentence 15? That sentence can be paraphrased as, 'Everyone who is not Pavel is owed money by Pavel.' This is a sentence structure we already know how to symbolise: 'For all x: if x is not Pavel, then x is owed money by Pavel.' In $\mathcal{L}_P$ with identity, this becomes $\forall x(x \neq p \rightarrow Opx)$.

In addition to sentences that use the word 'else', identity will be helpful when symbolizing some sentences that contain the words 'besides' and 'only.' Consider these examples:

17. No one besides Pavel owes money to Hikaru.
18. Only Pavel owes Hikaru money.

We add the constant h, which means Hikaru.

Sentence 17 can be paraphrased as, 'No one who is not Pavel owes money to Hikaru.' This can be translated as $\neg\exists x(x \neq p \wedge Oxh)$.

Sentence 18 can be paraphrased as, 'Pavel owes Hikaru *and* no one besides Pavel owes Hikaru money.' We have already translated one of the conjuncts, and the other is straightforward. Sentence 18 becomes $(Oph \wedge \neg\exists x(x \neq p \wedge Oxh))$.

6. Expressions of quantity

One of the reasons why it's useful to have an identity symbol, $=$, is that it allows us to say how many things there are of a particular kind. For example, consider these sentences:

19. There is at least one apple on the table.
20. There are at least two apples on the table.
21. There are at least three apples on the table.

Let the UoD be *things on the table*, and let A pick out the property of *being on the table*.

Sentence 19 does not require the use of the identity symbol. It can be translated adequately as $\exists x(Ax)$. That is, there is some apple on the table—perhaps many, but at least one.

It might be tempting to also translate Sentence 20 without identity. Yet consider the sentence $\exists x\exists y(Ax \wedge Ay)$. It means that there is some apple x in the UoD and some apple

y in the UoD. Since nothing precludes x and y from picking out the same member of the UoD, this would be true even if there were only one apple. In order to make sure that there are two *different* apples, we need an identity predicate. Sentence 20 needs to say that the two apples that exist are not identical, so it can be translated as $\exists x \exists y(Ax \wedge Ay \wedge x \neq y)$.

Sentence 21 requires talking about three different apples. It can be symbolised $\exists x \exists y \exists z(Ax \wedge Ay \wedge Az \wedge x \neq y \wedge y \neq z \wedge x \neq z)$. Continuing in this way, we could translate 'There are at least n apples on the table.'

Now consider these sentences:

22. There is at most one apple on the table.
23. There are at most two apples on the table.

Sentence 22 can be paraphrased as, 'It is not the case that there are at least *two* apples on the table.' This is just the negation of Sentence 20:

$$\neg \exists x \exists y(Ax \wedge Ay \wedge x \neq y)$$

Sentence 22 can also be approached in another way. It means that any apples that there are on the table must be the selfsame apple, so it can be translated as $\forall x \forall y\big((Ax \wedge Ay) \rightarrow x = y\big)$. The two translations are logically equivalent, so both are correct.

In a similar way, Sentence 23 can be translated in two equivalent ways. It can be paraphrased as, 'It is not the case that there are *three* or more distinct apples', so it can be translated as the negation of Sentence 21. Using universal quantifiers, it can also be translated as:

$$\forall x \forall y \forall z\big((Ax \wedge Ay \wedge Az) \rightarrow (x = y \vee x = z \vee y = z)\big).$$

The examples above are sentences about apples, but the logical structure of the sentences translates mathematical inequalities like $a \geq 3$, $a \leq 2$, and so on. We also want to be able to translate statements of equality which say exactly how many things there are. For example:

24. There is exactly one apple on the table.
25. There are exactly two apples on the table.

Sentence 24 can be paraphrased as, 'There is *at least* one apple on the table, and there is *at most* one apple on the table.' This is just the conjunction of Sentence 19 and Sentence 22 from above: $\big(\exists x(Ax) \wedge \forall x \forall y\big((Ax \wedge Ay) \rightarrow x = y\big)\big)$. This is a somewhat complicated way of going about it. It is perhaps more straightforward to paraphrase Sentence 24 as, 'There is a thing which is the only apple on the table.' Thought of in this way, the sentence can be translated $\exists x\big(Ax \wedge \neg \exists y(Ay \wedge x \neq y)\big)$.

Similarly, Sentence 25 may be paraphrased as, 'There are two different apples on the table, and these are the only apples on the table.' This can be translated as $\exists x \exists y\big(Ax \wedge Ay \wedge x \neq y \wedge \neg \exists z(Az \wedge x \neq z \wedge y \neq z)\big)$.

Finally, consider this sentence:

26. There are at most two things on the table.

It might be tempting to add a predicate so that T would mean ... is a thing on the table.' However, this is unnecessary. Since the UoD is the set of things on the table, all members of the UoD are on the table. If we want to talk about a *thing on the table*, we need only use a quantifier. Sentence 26 can be symbolised like Sentence 23 (which said that there were at most two apples), but leaving out the predicate entirely. That is, Sentence 26 can be translated as $\forall x \forall y \forall z(x = y \vee x = z \vee y = z)$.

7. Definite descriptions

Recall that a constant of $\mathcal{L}_P$ must refer to some member of the UoD. This constraint allows us to avoid the problem of non-referring terms (§6.1). Given a UoD that included only actually existing creatures but a constant c that meant 'chimera' (a mythical creature), sentences containing c would become impossible to evaluate.

The most widely influential solution to this problem was introduced by Bertrand Russell in 1905. Russell asked how we should understand this sentence:

27. The present king of France is bald.

The sub-phrase 'the present king of France' is supposed to pick out an individual by means of a definite description. However, there was no king of France in 1905 and there is none now. Since the description is a non-referring term, we cannot just define a constant f to mean 'the present king of France' and translate the sentence as Kf.

Russell's idea was that sentences that contain definite descriptions have a different logical structure than sentences that contain proper names, even though they share the same grammatical form. What do we mean when we use an unproblematic, referring description, like 'the highest peak in Washington state'? We mean that there is such a peak, because we could not talk about it otherwise. We also mean that it is the only such peak. If there was another peak in Washington state of exactly the same height as Mount Rainier, then Mount Rainier would not be *the* highest peak.

According to this analysis, Sentence 27 is saying three things. First, it makes an *existence* claim: there is some present king of France. Second, it makes a *uniqueness* claim: this guy is the only present king of France. Third, it makes a claim of *predication*: this guy is bald.

In order to symbolise definite descriptions in this way, we need the identity predicate. Without it, we could not translate the uniqueness claim which (according to Russell) is implicit in the definite description. Let the UoD be the set of *people actually living*, let F mean ... is the present king of France', and let B mean ... is bald.' Sentence 27 can then be translated as $\exists x(Fx \wedge \neg\exists y(Fy \wedge x \neq y) \wedge Bx)$. This says that *there is* someone who is the present king of France, he is the *only* present king of France, and he is bald. Understood in this way, Sentence 27 is meaningful but false. It says that a certain kind of person exists, but when no such person does.

The problem of non-referring terms is most vexing when we try to translate negations. So consider this sentence:

28. The present king of France is not bald.

According to Russell, this sentence is ambiguous. It could mean either of two things:

28a. It is not the case that the present king of France is bald.

28b. The present king of France is non-bald.

Both possible meanings negate sentence 27, but they put the negation in different places.

Sentence 28a is called a **wide-scope negation**, because it negates the entire sentence. It can be translated as $\neg\exists x(Fx \wedge \neg\exists y(Fy \wedge x \neq y) \wedge Bx)$. This does not say anything about the present king of France, but rather says that some sentence about the present king of France is false. Since Sentence 27 if false, Sentence 28a is true.

Sentence 28b says something about the present king of France. It says that he lacks the property of baldness. Like Sentence 27, it makes an existence claim and a uniqueness claim; it just denies the claim of predication. This is called **narrow-scope negation**. It can be translated as $\exists x(Fx \wedge \neg\exists y(Fy \wedge x \neq y) \wedge \neg Bx)$. Since there is no present king of France, this sentence is false.

Russell's theory of definite descriptions resolves the problem of non-referring terms and also explains why it seemed so paradoxical. Before we distinguished between the wide-scope and narrow-scope negations, it seemed that sentences like 28 should be both true and false. By showing that such sentences are ambiguous, Russell showed that they are true understood one way but false understood another way.

Practice Exercises

Part A Choosing your own UoD and symbolisation key, translate each English-language sentence into $\mathcal{L}_P$.

1. All the food is on the table.
2. If the food has not run out, then it is on the table.
3. Everyone likes some of the food.
4. If anyone likes the food, then Eli does.
5. Francesca only likes the dishes that have run out.
6. Francesca likes no one, and no one likes Francesca.
7. Eli likes anyone who likes some of the food.
8. Eli likes anyone who likes the people that he likes.
9. If there is a person on the table already, then all of the food must have run out.

⋆ **Part B** Using the symbolisation key given, translate each English-language sentence into $\mathcal{L}_P$.

UoD :	The set of all people
D :	. . . dances ballet
F :	. . . is female
M :	. . . is male
C :	. . . is a child of . . .
S :	. . . is a sibling of . . .
e :	Elmer
j :	Jane
p :	Patrick

1. All of Patrick's children are ballet dancers.
2. Jane is Patrick's daughter.
3. Patrick has a daughter.
4. Jane is an only child.
5. All of Patrick's daughters dance ballet.
6. Patrick has no sons.
7. Jane is Elmer's niece.
8. Patrick is Elmer's brother.
9. Patrick's brothers have no children.
10. Jane is an aunt.
11. Everyone who dances ballet has a sister who also dances ballet.
12. Every man who dances ballet is the child of someone who dances ballet.

Part C Using the symbolisation key given, translate each English-language sentence into $\mathcal{L}_P$ with identity. The last sentence is ambiguous and can be translated two ways; you should provide both translations. (Hint: Identity is only required for the last four sentences.)

UoD :	The set of all people
K :	. . . knows the combination to the safe
S :	. . . is a spy
V :	. . . is a vegetarian

T : ... trusts ...
h : Hofthor
i : Ingmar

1. Hofthor is a spy, but no vegetarian is a spy.
2. No one knows the combination to the safe unless Ingmar does.
3. No spy knows the combination to the safe.
4. Neither Hofthor nor Ingmar is a vegetarian.
5. Hofthor trusts a vegetarian.
6. Everyone who trusts Ingmar trusts a vegetarian.
7. Everyone who trusts Ingmar trusts someone who trusts a vegetarian.
8. Only Ingmar knows the combination to the safe.
9. Ingmar trusts Hofthor, but no one else.
10. The person who knows the combination to the safe is a vegetarian.
11. The person who knows the combination to the safe is not a spy.

$\star$ **Part D** Using the symbolisation key given, translate each English-language sentence into $\mathcal{L}_P$ with identity. The last two sentences are ambiguous and can be translated two ways; you should provide both translations for each.

UoD : The set of cards in a standard deck
B : ... is black
C : ... is a club
D : ... is a deuce
J : ... is a jack
M : ... is a man with an axe
O : ... is one-eyed
W : ... is wild

1. All clubs are black cards.
2. There are no wild cards.
3. There are at least two clubs.
4. There is more than one one-eyed jack.
5. There are at most two one-eyed jacks.
6. There are two black jacks.
7. There are four deuces.
8. The deuce of clubs is a black card.
9. One-eyed jacks and the man with the axe are wild.
10. If the deuce of clubs is wild, then there is exactly one wild card.
11. The man with the axe is not a jack.
12. The deuce of clubs is not the man with the axe.

Part E Using the symbolisation key given, translate each English-language sentence into $\mathcal{L}_P$ with identity. The last two sentences are ambiguous and can be translated two ways; you should provide both translations for each.

UoD : The set of animals in the world
B : ... is in Farmer Brown's field.
H : ... is a horse.
P : ... is a Pegasus.

W : ... has wings.

1. There are at least three horses in the world.
2. There are at least three animals in the world.
3. There is more than one horse in Farmer Brown's field.
4. There are three horses in Farmer Brown's field.
5. There is a single winged creature in Farmer Brown's field; any other creatures in the field must be wingless.
6. The Pegasus is a winged horse.
7. The animal in Farmer Brown's field is not a horse.
8. The horse in Farmer Brown's field does not have wings.

Part II

Probability

Chapter 7

Introduction to Probabilities

In this final part of the book, we're going to take a brief look at probabilities. Probabilities are important for many reasons, not least because we cannot do all our reasoning with deductive logic alone. Deductive logic works best when we're dealing with *certainties*: if an argument is valid and you're certain that its premises are true, then you can be certain that its conclusion is also true. But such situations are rare. More likely, you will be at least a little uncertain about most of the propositions you consider.

For reasoning with uncertain premises, we need a grasp of how probabilities work. Probability theory forms the backbone of many areas in the sciences, including physics, biology, and the social sciences, and is very important for many arguments in philosophy. In this chapter, we start with the basic theory of probabilities.

1. Background Concepts

Though there are some minor disagreements here and there regarding the *mathematical* theory of probability, overall there is a great deal of agreement on the subject. By contrast, there's far less agreement on *philosophical* theories of probability. Philosophers have long debated, and continue to debate, what it really means to say that some proposition is probable or improbable. We will not survey each of the many theories that have been put forward as to what probability statements mean. Instead, we're going to work with one interpretation that's very common (though not universal) in philosophy—that probabilities represent rational degrees of confidence.

Let's start with **degrees of confidence**. You are likely already familiar with the idea that **certainty** in the truth of some proposition corresponds to having 100% confidence in its truth. We speak in this sort of way regularly, in ordinary English. And you will also be familiar with the idea that we can have more or less confidence in the truth of some proposition. One might say, for example, that they're only about 80% confident that it will rain tomorrow, or 50% confident that there is intelligent extraterrestrial life. These are ways of expressing different degrees of confidence held towards a proposition. The idea is simply that we can represent how confident you are regarding the truth of some proposition—regarding how likely it is, or how probable it is that it's true—using a scale from 0% to 100%. The higher the degree, the more confident you are. At 0% confidence, you are completely certain the proposition in question is false; at 100%, you are completely certain the proposition in question is true; and for values between 0% and 100%, you have various different levels of confidence.

If we can have different degrees of confidence, then it seems those degrees may be either **rational** or **irrational**. For example, consider the proposition *if snow is white then snow is white*. Since that is a tautology, we can know that it must be true as a matter of logic alone. So it seems it would be rational to have 100% confidence in its truth; certainly, it would be irrational to be only 5% confident that it is true. It would likewise seem irrational to be certain that *water is wet and it is not wet*, since that is a logical contradiction and can never be true. In fact, any degree of confidence greater than 0% seems irrational for any logical contradictions. These examples indicate that there is a close relationship between logic and rational degrees of confidence. It is rational to be 100% confident in the truth of a tautology, because logic tells us a tautology must be true regardless of what the world is like. And it is rational to be 0% confident in the truth of a contradiction, because logic tells us a contradiction must be false regardless of what the world is like.

Here is another example. Suppose you are 70% confident that *it will rain tomorrow*. Would it then be rational to be 20% confident that *it will rain tomorrow or it will snow*? It seems not, because whenever the first proposition is true the second proposition must also be true, and the second proposition is also true in some cases where the first one isn't. So the probability of the second proposition must be *at least as great* as the probability of the first one, since it is true in a larger number of cases. Again, this example indicates that there is a close relationship between rational degrees of confidence and logic. In this case: if one proposition P logically entails another proposition Q, then your confidence in P should be no greater than your confidence in Q.

Representing probabilities

We will need a way of formalising our talk of probabilities. This means we will need a formal language for expressing propositions, as well as something to represent the probabilities that attach to those propositions. For these chapters, we will go back to using the simpler formal language of sentential logic that we developed in chapters 2 through 4.

Above, we talked about probabilities as percentages, taking values between 0% and 100% inclusive. But in the mathematical theory of probability, it is more common to represent probabilities as numbers between 0 and 1. You can translate from the 0-to-1 scale to the 0-to-100 scale simply by multiplying by 100, like so:

- ▷ $\mathcal{X}$ has a probability of 0.1 = $\mathcal{X}$ has a probability of 10%
- ▷ $\mathcal{X}$ has a probability of 0.25 = $\mathcal{X}$ has a probability of 25%
- ▷ $\mathcal{X}$ has a probability of 0.5 = $\mathcal{X}$ has a probability of 50%
- ▷ $\mathcal{X}$ has a probability of 0.6475 = $\mathcal{X}$ has a probability of 64.75%
- ▷ $\mathcal{X}$ has a probability of 1 = $\mathcal{X}$ has a probability of 100%

Next, we don't want to be writing "... has a probability of ... " over and again, as that would quickly become tedious. So instead, we'll represent the probability of a proposition using a **probability function**. We represent a probability function with $pr(\cdot)$. It takes a proposition as input (represented by the sentence that goes inside the brackets), and gives us a probability value between 0 and 1 inclusive as output. For example, if the probability of the proposition expressed by the sentence S is 0.25, and the probability of the proposition expressed by the sentence $S \vee W$ is 0.5, then we write:

$$pr(S) = 0.25$$
$$pr(S \vee W) = 0.5$$

Every proposition has exactly *one* probability value at any given time. Probabilities may change over time—for example, if you get new evidence that the proposition is true, which will change your rational degree of confidence regarding that proposition—but it doesn't make sense to say that the probability of some is, say, both 0.3 and 0.5 *at the same time.*

> A PROBABILITY is a number between 0 and 1 which attaches to whole propositions. '$pr(\mathcal{X}) = n$' means that the probability of the proposition expressed by the sentence $\mathcal{X}$ is the numerical value n.

2. The Probability Axioms

In mathematics, an **axiom** is a basic statement upon which a more complex mathematical structure can be constructed. In this section we're going to look at the three basic axioms of probability theory. As you'll soon see, they are really quite simple indeed—but from such small beginnings, we can do quite a lot!

The original version of these axioms were described by the Russian mathematician Andrey Kolmogorov, in his 1933 book *Foundations of the Theory of Probability.* Here, we'll work with a lightly simplified version of the axioms that foregoes some of the mathematical details.

> **Axiom 1:** Non-Negativity
> For any $\mathcal{X}$, $pr(\mathcal{X}) \geq 0$

That is, the probability of any proposition whatsoever can never be less than 0. This should be intuitive—it doesn't make sense to say that the probability of something's being true is *less than* 0%! Axiom 1 is usually called the **non-negativity** axiom, because it tells us that probability values are never negative.

> **Axiom 2:** Normalisation
> If $\mathcal{X}$ is a tautology then $pr(\mathcal{X}) = 1$

A tautology is a proposition which must be true as a matter of logic, regardless of how the world turns out to be. We can know that each of these is true without doing any empirical work, purely by our logical reasoning. Axiom 2 tells us that the probability of any tautology is always 1. It is usually called the **normalisation** axiom because it has the effect of setting an upper limit of 1 for the probability of any proposition—nothing can be more probable than a tautology.

> **Axiom 3:** Additivity
> If $\mathcal{X}$ and $\mathcal{Y}$ are jointly inconsistent then $pr(\mathcal{X} \vee \mathcal{Y}) = pr(\mathcal{X}) + pr(\mathcal{Y})$

Recall from §4.4 that two propositions are jointly inconsistent whenever it's not possible for both to be true at the same time. So, Axiom 3 says if $\mathcal{X}$ and $\mathcal{Y}$ are *jointly inconsistent*, then the probability of their disjunction $\mathcal{X} \vee \mathcal{Y}$ is equal to the sum of the probabilities of $\mathcal{X}$ and $\mathcal{Y}$ individually. Axiom 3 is called the **additivity** axiom.

And that's it. Those three axioms will be the basis of everything we discuss in this chapter.

3. Basic Laws of Probability

Let's apply the three probability axioms to derive some **laws of probability**. A law of probability is any statement about probabilities that follows logically from the three axioms of probability.

The probabilities of $\mathcal{X}$ and $\neg\mathcal{X}$ always sum to 1

Suppose we toss a fair 6-sided die, with sides labelled '1' through '6'. We assume that the die *must* land one one side or another; it cannot land on multiple sides at once, it cannot land on an edge, and it cannot stop in midair. When we say that the coin is fair, we mean that every side has the same probability, 1/6, of showing up as every other side. There are therefore 6 possible outcomes, which we can symbolise like so:

D_1 : The die will land on side 1
D_2 : The die will land on side 2
D_3 : The die will land on side 3
D_4 : The die will land on side 4
D_5 : The die will land on side 5
D_6 : The die will land on side 6

We can also create more complex propositions out of these sentence letters. For instance, $D_1 \vee \neg D_2$ is the proposition *the die either lands on side 1 or not side 2*, while $D_2 \vee D_4 \vee D_6$ is equivalent to *the die lands on an even side.*

Consider first the proposition expressed by $D_1 \vee \neg D_1$; i.e., *the coin will land on side 1, or it will not land on side 1.* Since this is a tautology, we know by the normalisation axiom that it has a probability of 1:

$$pr(D_1 \vee \neg D_1) = 1$$

But note that D_1 and $\neg D_1$ are jointly inconsistent. Therefore, according to the additivity axiom,

$$pr(D_1 \vee \neg D_1) = pr(D_1) + pr(\neg D_1)$$

Putting those two together, we know that the the probabilities of D_1 and $\neg D_1$ must sum to 1. Since the probability of the die landing on side 1 is (by the stipulation above) equal to 1/6, it follows that the probability of its landing on some other side is 5/6. Similar reasoning will apply to *any* sentence of the form $\mathcal{X} \vee \neg\mathcal{X}$. Hence we have our first law of probability, which follows from the normalisation and additivity axioms:

Law 1: The probabilities of $\mathcal{X}$ and $\neg\mathcal{X}$ always sum to 1.
$pr(\mathcal{X}) + pr(\neg\mathcal{X}) \ = \ 1$
Equivalently: $pr(\mathcal{X}) \ = \ 1 - pr(\neg\mathcal{X})$

The probability of a contradiction is always zero

Note that our first law of probability also implies that the probability of a contradiction must always be zero. The law tells us that for *any* sentence $\mathcal{X}$, $pr(\mathcal{X}) \ = \ 1 - pr(\neg\mathcal{X})$. That includes tautological sentences, such as $D_1 \vee \neg D_1$:

$$pr(D_1 \vee \neg D_1) \ = \ 1 - pr(\neg(D_1 \vee \neg D_1))$$

Since we already know that $pr(D_1 \vee \neg D_1) = 1$, it follows that the probability of the contradiction, $\neg(D_1 \vee \neg D_1)$, must be zero. Again, the same kind of reasoning applies to all contradictions; hence we have our second law:

Law 2: The probability of a contradiction is always zero.
If $\mathcal{X}$ is a contradiction then $pr(\mathcal{X}) = 0$

Additivity for more than two propositions

Next consider the proposition *the coin will land on an odd side*, which can symbolised as $D_1 \vee D_3 \vee D_5$. This is not a tautology, but note that D_1, D_3 and D_5 are all inconsistent with one another. Therefore the additivity axiom tells us:

$$pr(D_1 \vee D_3) = pr(D_1) + pr(D_3)$$
$$pr(D_1 \vee D_5) = pr(D_1) + pr(D_5)$$
$$pr(D_3 \vee D_5) = pr(D_3) + pr(D_5)$$

But it tells us something further as well. If D_1 is inconsistent with D_3 and with D_5 separately, then it is also inconsistent with their disjunction $D_3 \vee D_5$. After all, $D_3 \vee D_5$ is true if and only if either D_3 is true or D_5 is true, but neither can be true if D_1 is true. Consequently, the additivity axiom tells us:

$$pr(D_1 \vee D_3 \vee D_5) = pr(D_1) + pr(D_3 \vee D_5)$$

But since D_3 and D_5 are jointly inconsistent,

$$pr(D_3 \vee D_5) = pr(D_3) + pr(D_5)$$

Therefore, from the equation above, we know:

$$pr(D_1 \vee D_3 \vee D_5) = pr(D_1) + pr(D_3) + pr(D_5)$$

Since we've said that D_1, D_3 and D_5 all have the same probability of 1/6, the probability of $D_1 \vee D_3 \vee D_5$ must be 3/6, or 50%. This should be as expected: with a fair 6-sided die, the probability of landing on an odd number is exactly half. The same sort of thing is true any time we have any (finite) collection of propositions which are all inconsistent with one another, *even if they don't have all the same probability*. For example, if $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are all pairwise inconsistent with one another, then:

$$pr(\mathcal{X} \vee \mathcal{Y} \vee \mathcal{Y}) = pr(\mathcal{X}) + pr(\mathcal{Y}) + pr(\mathcal{Z})$$

Hence, we have our third law of probability, which generalises the additivity axiom:

Law 3: Additivity for more than two propositions.
If it is not possible for more than one member of a (finite) set of propositions $\{\mathcal{X}, \mathcal{Y}, \ldots, \mathcal{Z}\}$ to be true at the same time, then:

$$pr(\mathcal{X} \vee \mathcal{Y} \vee \ldots \vee \mathcal{Z}) \; = \; pr(\mathcal{X}) + pr(\mathcal{Y}) + \ldots + pr(\mathcal{Z})$$

This law also tells us something about **partitions**. A partition is any set of propositions that satisfies both of the following conditions:

1. It is not possible for every proposition in the set to be false
2. It is not possible for more than proposition in the set to be true

Sometimes, the propositions that form a partition are said to be **mutually exclusive and jointly exhaustive**. The 'mutually exclusive' means that each proposition in the set is jointly inconsistent with every other proposition in the set. The 'jointly exhaustive' means that the propositions in the set cover every possibility, with nothing left over. In short, whenever you have a partition, exactly one member of the partition can and must be true.

Since we have assumed above that the die must land on one of its six sides, we can consider the set $\{D_1, D_2, D_3, D_4, D_5, D_6\}$ to be a partition of proposition. (It is not the only one; for example, $\{D_1, \neg D_1\}$ is also a partition, as is $\{D_1 \vee D_2 \vee D_3, D_4 \vee D_5, D_6\}$, and many more.) Now since at least one of the propositions in a partition must be true, the *disjunction* of every proposition in a partition must always have a probability of 1:

$$pr(D_1 \vee D_2 \vee D_3 \vee D_4 \vee D_5 \vee D_6) = 1$$

But since it's not possible for more than one proposition in a partition to be true, the second law of probability tells us:

$$pr(D_1) + pr(D_2) + pr(D_3) + pr(D_4) + pr(D_5) + pr(D_6) = 1$$

More generally, any time you have a partition of propositions, the probabilities of all the propositions in that partition must always sum to one.

Equivalent propositions always have the same probability

The next law of probability is very useful. Suppose that two sentences are logically equivalent. For example, $\neg D_1 \vee \neg D_2$ is logically equivalent to $\neg(D_2 \wedge D_1)$. Then, we should expect that the probability of $\neg D_1 \vee \neg D_2$ is exactly equal to the probability of $\neg(D_2 \wedge D_1)$. Since logic tells us that they are true under exactly the same circumstances, it would be absurd to consider one to be highly probable while considering the other to be not very probable. So we should want the following law of probability to be true:

> **Law 4:** Equivalent propositions always have the same probability.
> If $\mathcal{X}$ and $\mathcal{Y}$ are logically equivalent, then $pr(\mathcal{X}) = pr(\mathcal{Y})$

We can derive this from the laws above, using a bit of logic and algebra. The reasoning goes as follows. If $\mathcal{X}$ and $\mathcal{Y}$ are logically equivalent, then $\neg\mathcal{X}$ and $\neg\mathcal{Y}$ must also be logically equivalent. (If two sentences are true in exactly the same conditions, then they must also be false in exactly the same conditions.) Consequently, if $\mathcal{X}$ and $\mathcal{Y}$ are logically equivalent, then $\mathcal{X} \vee \neg\mathcal{Y}$ must be a tautology, and $\mathcal{X}$ is inconsistent with $\neg\mathcal{Y}$. Hence:

$$pr(\mathcal{X}) + pr(\neg\mathcal{Y}) = 1$$

But we also know that:

$$pr(\mathcal{Y}) + pr(\neg\mathcal{Y}) = 1$$

Therefore, it must be the case that:

$$pr(\mathcal{X}) = pr(\mathcal{Y})$$

General law for probability of disjunction

The final law of probability that we'll look at in this chapter is also going to be another generalisation of the additivity axiom. So far, we know how to work out the probabilities of disjunctions of propositions when those propositions are inconsistent with one another. But what if we want to know the probability of the disjunction of a pair of propositions that *can* both be true at the same time? For this, we can apply the following law:

Law 5: General law for probability of disjunction.
$pr(\mathcal{X} \vee \mathcal{Y}) = pr(\mathcal{X}) + pr(\mathcal{Y}) - pr(\mathcal{X} \wedge \mathcal{Y})$

Suppose that we want to know the probability that *the die will either land on something other than side 4 or on side 1.* In other words, we want to know the value of $pr(\neg D_4 \vee D_1)$. Since the die is fair we know that $pr(D_1) = 1/6$. From Law 1 we know that $pr(\neg D_4) = 5/6$. However, if we simply add $pr(D_1)$ to $pr(\neg D_4)$, we would get 1—and that is the wrong result. Why? Because $\neg D_4 \vee D_1$ would be false if the die lands on side 4, and there is still a 1/6 probability that it will land on that side. So $pr(\neg D_4 \vee D_1)$ should be less than 1.

The problem is that when we add $pr(D_1)$ to $pr(\neg D_4)$, we are double-counting the event where *both* $pr(\neg D_4)$ and $pr(D_1)$ are true. To rectify the situation, we need to subtract the case where they are both true—we need to subtract $pr(\neg D_4 \wedge D_1)$. So:

$$pr(\neg D_4 \vee D_1) = pr(\neg D_4) + pr(D_1) - pr(\neg D_4 \wedge D_1)$$

In this case, the probability of the conjunction $\neg D_4 \wedge D_1$ is just the probability of D_1, because $\neg D_4 \wedge D_1$ is true if and only if the die lands on side 1. Hence,

$$pr(\neg D_4 \vee D_1) \ = \ 5/6 + 1/6 - 1/6 \ = 5/6$$

This is the right result: the probability of $\neg D_4 \vee D_1$ is just the probability that the die will land on side 1, 2, 3, 5 or 6, which is 5/6.

The additivity axiom applies only to disjunctions where the disjuncts are inconsistent with one another. In that case, there is no need to worry about double-counting, since the propositions can never be true at the same time. Law 5 applies to any disjunctions whatsoever, regardless of whether the disjuncts are mutually exclusive. If $\mathcal{X}$ and $\mathcal{Y}$ are inconsistent with one another, then the probability of their conjunction is zero, so in that case the additivity axiom and Law 5 say the very same thing.

Chapter 6: Key ideas

- In philosophy, statements about probability are often (though not always) interpreted as statements about **degrees of confidence**. Roughly, to say that some proposition P has a probability of, say, 50%, is to say that one is 50% confident of its being true. The laws of probability are then interpreted as laws of rationality—they describe what your degrees of confidence ought to be like if you are fully rational. There is a very close connection between rational degrees of confidence and logic.

- Claims about probabilities are usually represented with a **probability function**, $pr(\cdot)$, and probabilities values are represented using numbers between 0 and 1 inclusive. We read '$pr(\mathcal{X}) = n$' as saying that the probability of the proposition expressed by the sentence $\mathcal{X}$ is the value n.

- There are three **axioms of probability**. These are basic laws from which other laws of probability can be derived. The three axioms are:

> **Axiom 1:** Non-Negativity
> For any $\mathcal{X}$, $pr(\mathcal{X}) \geq 0$
>
> **Axiom 2:** Normalisation
> For any $\mathcal{X}$, if $\mathcal{X}$ is a tautology then $pr(\mathcal{X}) = 1$
>
> **Axiom 3:** Additivity
> For any $\mathcal{X}$ and $\mathcal{Y}$, if $\mathcal{X}$ and $\mathcal{Y}$ are jointly inconsistent then $pr(\mathcal{X} \vee \mathcal{Y}) = pr(\mathcal{X}) + pr(\mathcal{Y})$

- There are infinitely many laws of probability. Some of the most commonly used are:

> **Law 1**: The probabilities of $\mathcal{X}$ and $\neg\mathcal{X}$ always sum to 1.
> $pr(\mathcal{X}) + pr(\neg\mathcal{X}) \ = \ 1$
> Equivalently: $pr(\mathcal{X}) \ = \ 1 - pr(\neg\mathcal{X})$
>
> **Law 2:** The probability of a contradiction is always zero.
> If $\mathcal{X}$ is a contradiction then $pr(\mathcal{X}) = 0$
>
> **Law 3:** Additivity for more than two propositions.
> If it is not possible for more than one member of a (finite) set of propositions $\{\mathcal{X}, \mathcal{Y}, \ldots, \mathcal{Z}\}$ to be true at the same time, then:
>
> $$pr(\mathcal{X} \vee \mathcal{Y} \vee \ldots \vee \mathcal{Z}) \ = \ pr(\mathcal{X}) + pr(\mathcal{Y}) + \ldots + pr(\mathcal{Z})$$
>
> **Law 4:** Equivalent propositions always have the same probability.
> If $\mathcal{X}$ and $\mathcal{Y}$ are logically equivalent, then $pr(\mathcal{X}) = pr(\mathcal{Y})$
>
> **Law 5:** General law for probability of disjunction.
> $pr(\mathcal{X} \vee \mathcal{Y}) = pr(\mathcal{X}) + pr(\mathcal{Y}) - pr(\mathcal{X} \wedge \mathcal{Y})$

Practice Exercises

Part A
Suppose you have a standard deck of 52 cards. What is:

1. The probability that the card is hearts?
2. The probability that the card is not a jack?
3. The probability that the card is hearts or clubs?
4. The probability that the card is an ace or a king?
5. The probability that the card is hearts or an ace?
6. The probability that the card is hearts and not an ace?
7. The probability that the card is an ace or a red queen?
8. The probability that the card is either not hearts, or an ace?

⋆ **Part B**
Suppose that $pr(S) = 0.25$, $pr(V) = 0.25$, and $pr(S \wedge V) = 0.125$.

1. Are S and V consistent with one another?
2. Are S and V logically equivalent to one another?
3. What is the value of $pr(S \vee V)$?
4. What is the value of $pr((S \vee V) \wedge S)$?

Part C
In addition to providing an answer for the following questions, provide an explanation of why the answer is correct (you might find it helpful to cite one or more of the laws and axioms, and how it follows.)

1. Suppose that $pr(A) = 0.5$, $pr(B) = 0.25$, and $pr(A \vee B) = 0.75$. What is the value of $pr(A \wedge B)$?
2. Suppose that $pr(A) = 0.5$, $pr(B) = 0.25$, and $pr(A \vee B) = 0.75$. What is the value of $pr(A \wedge (B \vee \neg B))$?
3. Suppose that $pr(A) = 0.5$, $pr(B) = 0.5$, and $pr(A \vee B) = 0.5$. What is the value of $pr(A \wedge B)$?
4. Suppose that $pr(A) = 0.5$, $pr(B) = 0.5$, and $pr(A \vee B) = 0.5$. What is the value of $pr(A \vee (B \vee \neg B))$?
5. Suppose that $pr(A \vee B) = 0.5$, $pr(B) = 0.25$, $pr(A \wedge B) = 0.25$. What is the value of $pr(\neg A)$? $=$
6. Suppose that $pr(A \vee B) = 0.5$, $pr(B) = 0.25$, $pr(A \wedge B) = 0.25$. What is the value of $pr(\neg A \to B)$?
7. Suppose that A and B are mutually exclusive. Is it possible to have $pr(A) = 0.5$ and $pr(B) = 0.75$ at the same time?

Part D
Explain how one might try to derive the following additional laws of probability using the three axioms of this chapter, which hold for any $\mathcal{X}$ and $\mathcal{Y}$.

1. $pr(\mathcal{X} \wedge \mathcal{Y}) \leq pr(\mathcal{X})$
2. $pr(\mathcal{X}) \leq pr(\mathcal{X} \vee \mathcal{Y})$
3. $pr(\mathcal{X}) = pr(\mathcal{X} \vee \mathcal{Y})$ only if $pr(\mathcal{Y}) = pr(\mathcal{X} \wedge \mathcal{Y})$

Chapter 8

Conditional Probabilities

In Chapter 7, we dealt with unconditional probability statements. These are statements like 'The probability of that the coin will land tails is 50%'. However, when we're thinking probabilistically, we often have to deal with **conditional probabilities**. This chapter will develop the theory of conditional probability, and highlight some very important characteristics of good probabilistic reasoning.

1. Conditional versus Unconditional Probability

Suppose you come across a coin, and decide to toss it. Knowing nothing much about the coin, you probably expect it to be just like every other coin you've seen. In particular, you figure that there's a roughly 50% chance that when you toss it, the coin will land heads, and a roughly 50% chance that it will land tails. These are **unconditional probability** values. On the other hand, what is the probability that a coin will land heads *given that* it is biased 70% towards tails? In this case, the correct answer would be 0.3, or 30%. This value isn't the probability that the coin will land heads, because you may not *know* whether the coin is biased towards tails or not. Instead, it's the probability that you would say the proposition has, *if* you suppose that the coin were biased in such a way. Conditional probabilities are probabilities *under a condition.*

Likewise, we could ask: what is the probability that the coin will land heads *given that* it is biased 80% towards heads? In this case, the correct answer would be 0.8. We could even ask: what is the probability that the coin will land heads *given that* it lands heads. Here, the correct answer is 1.

When we want to say that the probability of one proposition (expressed by the sentence $\mathcal{X}$) under a particular condition (expressed by the sentence $\mathcal{Y}$) is some value n, we write '$pr(\mathcal{X}|\mathcal{Y}) = n$'. You can read this as 'The probability that $\mathcal{X}$ is true, *given that* $\mathcal{Y}$ is true, is n.' By way of example, suppose you have an ordinary, well-shuffled deck of 52 cards. You select a card at random from the deck. What is the unconditional probability that it is a heart? The answer is 1/4. That is,

$$pr(\text{The card is a heart}) = 1/4$$

On the other hand, what is the probability that the card is a heart *given that* it is a red card? This time, the answer is 1/2. So,

$$pr(\text{The card is a heart} \mid \text{The card is red}) = 1/2$$

Alternatively, what is the probability that the card is a heart, *given that* the card is black? The answer is 0. Hence,

pr(The card is a heart | The card is black) = 0

In these examples, the conditional probability of 'The card is a heart' is different than its unconditional probability depending on what the relevant condition is. Different conditions led to different conditional probabilities.

2. Independence and the Gambler's fallacy

What if we asked about the probability that the card is a heart, *given that* the cards in the deck have rounded rather than sharp corners? Presumably, the sharpness of the corners makes no difference to the probability that the card is a heart, so:

pr(The card is a heart) = 1/4
pr(The card is a heart | The card's corners are rounded) = 1/4

In the jargon of probability theory, we say in this case that whether the card is a heart is **probabilistically independent** (or just 'independent') of the roundness of the card's corners. If two propositions are not probabilistically independent of one another, then we say that they are **probabilistically correlated**. Generally:

> $\mathcal{X}$ is PROBABILISTICALLY INDEPENDENT of $\mathcal{Y}$ just in case $pr(\mathcal{X}) = pr(\mathcal{X}|\mathcal{Y})$; otherwise, $\mathcal{X}$ and $\mathcal{Y}$ are PROBABILISTICALLY CORRELATED.

Essentially, $\mathcal{X}$ is independent of $\mathcal{Y}$ if assuming $\mathcal{Y}$ makes no difference to the probability of $\mathcal{X}$. There are plenty of things which are probabilistically independent of whether a particular card is a heart: the number of dust motes on Mars, the day of the week that Christmas will be on this year, the average lifespan of a porcupine, and so on. On the other hand, whether the card is red is probabilistically correlated with whether the card is a heart, since assuming the that the card is red makes a difference to the probability that it is a heart.

Many very important results in the mathematics of probabilities depend on this notion of independence, and it is an extremely useful notion in the sciences that deal with statistics. However, people are apt to forget about independence when reasoning with probabilities. This forgetfulness is the basis of what's known as the **Gambler's fallacy**.

Suppose you have a fair coin—you know it's fair—and you decide to start tossing it over and over again. The first time it lands, it comes up heads. It had a 50% chance of coming up heads, so nothing strange there. Then it lands heads again, and again, and again. The coin has now landed heads four times in a row. The chances of this happening are quite low. You decide to flip it one more time. What are the chances that it will land heads this time? You might be tempted by the following thought:

> "The chances of getting five heads in a row are very small, so the next toss will probably come up tails. The coin is supposed to be fair, so I should be seeing about as many heads come up as tails—at this point, I'm *due* for a tails to show up."

The problem with this is that the outcome of each successive coin toss is probabilistically independent of the outcome of any previous toss, since there is no way for the outcome of any one toss to affect the outcome of any later toss. It's not as if the coin can *remember* how it landed on previous tosses, and tries to keep things fair. So:

$$\begin{aligned} pr(\text{The coin will land heads}) &= 1/2 \\ pr(\text{The coin will land heads}|\text{ The coin landed heads before}) &= 1/2 \end{aligned}$$

This may seem like an obvious mistake in retrospect, but failing to account for independence is very common. For example, at the roulette wheel people are more likely to vote red if the ball has landed black a few times in a row—as if the ball knew what it was doing, and wanted to keep things fair. And many slot machine users get the sense that their slot machine is more likely to hit the jackpot *this time* if they've had a previous bad run without any winnings. And the fallacy doesn't just arise with gamblers. For example, the sex of a newborn child is independent of the sex of any siblings it may have that have already been born. Nevertheless, when parents have (say) three boys in a row, they may start to think they are 'due' for a girl.

3. Conditional Probability and Conjunctions

We've now introduced unconditional and conditional probabilities, but we haven't said much about how they're related to one another. Here, we'll introduce one more axiom to the three axioms introduced in Chapter 6, which will allow us to connect conditional and unconditional probabilities to one another:

Axiom 4: Probability of a conjunction
$pr(\mathcal{X} \wedge \mathcal{Y}) = pr(\mathcal{X}\|\mathcal{Y}) \times pr(\mathcal{Y})$

To return to the cards example, what is the probability that a card drawn at random is both *red* and a *king*? Use the following symbolisation key:

R : The card is red
K : The card is a king

We can then symbolise 'The card is a red king' as $R \wedge K$. Since there are two red king in a deck of 52, $pr(R \wedge K)$ is $2/52 = 1/26$. That is the result we want, and that is what Axiom 4 tells us as well. It says:

$$pr(R \wedge K) = pr(R|K) \times pr(K)$$

In a standard deck, there are two red king and two black king, so the probability that the card is red *given that* it is a king is $2/4 = 1/2$. And the probability that the card is a king is $1/13$. When we multiply $1/2$ and $1/13$ together, we get $1/26$—the right result.

One more example. What is the probability of $\neg R \wedge \neg K$, that the card is not red and not a king? In this case, Axiom 4 says:

$$pr(\neg R \wedge \neg K) = pr(\neg R|\neg K) \times pr(\neg K)$$

The probability that the card is not red given that it is not a king is 1/2, and the probability that the card is not a king is 12/13. If we multiply these together, we get 6/13, or 24/52. This is the expected result: there are 26 cards that are not red, and two of them are king, so there are 24 cards out of 52 that are neither red nor a king.

We can now develop some new laws of probability. This next one is particularly useful, and follows immediately by rearranging the terms in Axiom 4:

Law 6: Definition of conditional probability.
If $pr(\mathcal{Y}) > 0$, then

$$pr(\mathcal{X}|\mathcal{Y}) = \frac{pr(\mathcal{X} \wedge \mathcal{Y})}{pr(\mathcal{Y})}$$

Next, recall the definition of probabilistic independence above: $\mathcal{X}$ is probabilistically independent of $\mathcal{Y}$ whenever $pr(\mathcal{X}) = pr(\mathcal{X}|\mathcal{Y})$. This plus Axiom 4 lets us derive another very useful law of probability:

Law 7: Probabilities of conjunctions with independent conjuncts.
If $\mathcal{X}$ and $\mathcal{Y}$ are probabilistically independent, $pr(\mathcal{X} \wedge \mathcal{Y}) = pr(\mathcal{X}) \times pr(\mathcal{Y})$

We've already seen an instance of this law in action: the probability that the card is red is probabilistically independent of whether it is a king, so:

$$pr(R|K) = pr(R) = 1/2$$

So, by Law 7,

$$pr(R \wedge K) = pr(R) \times pr(K)$$
$$pr(R \wedge K) = 1/2 \times 1/13$$
$$pr(R \wedge K) = 1/26$$

This is exactly the right result, and it's much a easier formula to use for working out the probability of a conjunction than Axiom 4. But be careful! the probability of a conjunction is only equal to the product of the probabilities of its conjuncts *only if* those conjuncts are probabilistically independent of one another. It is quite common for students to forget the independence condition, and assume in order to calculate the probability of a conjunction one simply needs to multiply the probabilities of the conjunctions. That only applies when the conjuncts are not probabilistically correlated—if they are, then you must use Axiom 4 to work out the probability of a conjunction instead.

4. Bayes' Theorem and the Base Rate Fallacy

We are now in a position to introduce one of the most important equations in the theory of probability, **Bayes' theorem**. A useful formulation of the theorem is this:

Law 8: Bayes' Theorem
If $pr(\mathcal{Y}) > 0$, then:

$$pr(\mathcal{X}|\mathcal{Y}) = \frac{pr(\mathcal{Y}|\mathcal{X}) \times pr(\mathcal{X})}{pr(\mathcal{Y})}$$

Included here is a proof, for those who are interested in how Bayes' theorem follows from the axioms above. It is relatively simple, but you should feel free to skip over it if you prefer, or come back to it later:

> Axiom 4 says that:
>
> $$pr(\mathcal{X} \wedge \mathcal{Y}) = pr(\mathcal{X}|\mathcal{Y}) \times pr(\mathcal{Y})$$
>
> Axiom 4 also says that:
>
> $$pr(\mathcal{Y} \wedge \mathcal{X}) = pr(\mathcal{Y}|\mathcal{X}) \times pr(\mathcal{X})$$
>
> Since $\mathcal{X} \wedge \mathcal{Y}$ and $\mathcal{Y} \wedge \mathcal{X}$ are logically equivalent, we know that:
>
> $$pr(\mathcal{X} \wedge \mathcal{Y}) = pr(\mathcal{Y} \wedge \mathcal{X})$$
>
> Therefore:
>
> $$pr(\mathcal{X}|\mathcal{Y}) \times pr(\mathcal{Y}) = pr(\mathcal{Y}|\mathcal{X}) \times pr(\mathcal{X})$$
>
> Now, provided that $pr(\mathcal{Y}) > 0$, we can divide both sides by $pr(\mathcal{Y})$ to get:
>
> $$pr(\mathcal{X}|\mathcal{Y}) = \frac{pr(\mathcal{Y}|\mathcal{X}) \times pr(\mathcal{X})}{pr(\mathcal{Y})}$$

Bayes' theorem tells us that the conditional probability $pr(\mathcal{X}|\mathcal{Y})$ depends on three factors (assuming $pr(\mathcal{Y}) > 0$):

1. The unconditional probability of $\mathcal{X}$,
2. The probability of $\mathcal{Y}$ conditional on $\mathcal{X}$, and
3. The unconditional probability of $\mathcal{Y}$.

Why is this important? Because it turns out that we humans are often quite bad at reasoning in accordance with Bayes' theorem—we often fail to consider all three factors.

Here is an example. Suppose there is a terrible disease going around, we'll call it COVID-23. This disease has no particular outward symptoms, but after three days it kills you. You know that 1 in 1000 people currently have COVID-23, but you don't know if you currently have it. Luckily, scientists have developed a test. If you have the disease, then the test *will* tell you that you have it with 100% accuracy. In the jargon, we say that the test has 0% probability of returning a **false negative**—the probability that the test will say you have COVID-23 when you don't is exactly 0. However, 5% of the time when someone *doesn't* have COVID-23, the test says that they do. In this case, the test has a 5% probability of returning a **false positive**. The disease initially shows no symptoms, and you have no particular reason to think that you're more likely to have the disease than anyone else. Nevertheless, you are worried about your health, so you go in for a test. It comes back positive.

Given that, what's the probability that you have COVID-23, given the positive result? Most people will say that the probability you have COVID-23 in this case is very high—many will say that it is 95%. This is an instance of what's usually called the **Base Rate Fallacy**. The actual probability is closer to 2%.

This is the result we get when we use Bayes' theorem. Use this symbolisation key:

C : You have COVID-23
P : The test result is positive

What we want to know is the probability of C given P, the value of $pr(C|P)$. Bayes' theorem tells us we need three values:

1. The unconditional probability of C,
2. The probability of P conditional on C, and
3. The unconditional probability of P.

Since the background prevalence of the disease is 1/1000, the unconditional probability that you have COVID-23 is 1/1000. We also know that the probability of P conditional on C is 1, because the test always returns a positive result for people who actually have the disease. So according to Bayes' theorem,

$$pr(C|P) = \frac{pr(C) \times pr(P|C)}{pr(P)}$$

Which we now know is:

$$pr(C|P) = \frac{0.001 \times 1}{pr(P)} = \frac{0.001}{pr(P)}$$

The only thing we need now is the unconditional probability of P. This is the probability that the test will return a positive result *regardless* of whether you have COVID-23 or not. We can easily work this out. In the population of 100,000, we've said that the test will return a true positive for 100 people and a false positive for 4995 people. So, it will return a positive result for 5095 people out of 100,000 in total. So:

$$pr(P) = 5095/100000 = 0.05095$$

We can then work out the value of $pr(C|P)$:

$$pr(C|P) = \frac{0.01}{pr(P)} = \frac{0.001}{0.05095} \approx 0.02$$

In other words, even given the highly accurate testing procedure, the probability you have the disease given the positive test result is just under 2%.

It can be more intuitive to think about this result in terms of proportions. Imagine that everyone in a town of 100,000 people all decided to take this test. Since the background prevalence of the disease is 1 in 1000, we should expect that the test will correctly return a positive result for all 100 people who have COVID-23. However, of the remaining 99900 people in the population—all of whom do not have the disease—the test will *also* incorrectly say that 5% of *them* have COVID-23 as well. In other words, the test will return a false positive for 4995 people. So if you receive a positive test result, there are two possibilities—you might be one of the 100 people who actually have COVID-23, or you might be one of the 4995 people who do not have it. So it's about 50 times more likely that you do not have the disease, even given the positive result.

The base rate fallacy arises when we fail to consider the unconditional probability of the evidence we are using to form our judgements. In the example just given, when we learn that the test result is positive, we focus on the fact that the test will always give a positive result for someone who has the disease and will only give a positive result for someone who doesn't have the disease 5% of the time. We therefore consider it very likely that we have the disease if the test has a positive result. In so doing, though, we didn't consider the fact that the number of people who *don't* have the disease is vastly larger than the number of people who *do*, and therefore that 'small' 5% probability of a false positive actually translates into a very large number of false positives in comparison to the much smaller number of true positives.

The base rate fallacy is extremely important, and unfortunately very common in everyday reasoning. As the example shows, committing the fallacy can lead us to over-estimate the importance of test results. This is obviously problematic in situations where a correct understanding of the test results is crucial. In medical situations it can lead us to over-estimate the likelihood of a disease being present, leading to a misdiagnosis (with potentially life-threatening consequences). Similarly, in legal situations the base rate fallacy can cause us to wrongly judge the importance of a piece of evidence or testimony. Indeed, sometimes the base rate fallacy is also known as the **prosecutor's fallacy**, after hypothetical unscrupulous prosecutors who may try to use it to make their evidence appear stronger than it really is.

Here is one more example of Bayes' theorem in action. It is illegal to drink and drive, and police use breathalyzers to determine whether drivers are over the legal limit. However, no breathalyzer is perfect, and it's important to take this into account when we're considering the importance of breathalyzer results. Suppose that the breathalyzer correctly picks up on drunkenness 99% of the time, and only returns a false positive 1% of the time. Suppose that someone—let's call him 'Fred'—has been pulled over by police one night and given a breathalyzer test, and it indicates drunkenness. So, what is the probability that Fred was drunk driving *given that* the breathalyzer indicates that he was?

Use the following key:

D : Fred was drunk driving
T : The breathalyzer indicates drunkenness

We know from the description above that:

$$pr(T|D) = 0.99$$
$$pr(T|\neg D) = 0.01$$

So what is the value of $pr(D|T)$? By now you will know better than to answer '99%,' which would be the common but fallacious answer. As it turns out, we haven't yet been given enough information to determine what $pr(D|T)$ actually is. This is because we don't yet know the base rate. That is, we do not know the unconditional probability that Fred was drink driving, $pr(D)$. If we knew that value, then we could work out the value of $pr(D|T)$.

Suppose that only 1 in every 100 drivers are drunk driving on this particular night. This gives us a value for the base rate, $pr(D)$. For a population of 10,000 drivers, there will be 100 drunks on the road. If every one of them were tested, the breathalyzer test would catch 99 of them. However, the test would also incorrectly show that 99 out of the 9,900 *non*-drunk drivers were driving drunk. So, of the 198 people in total for whom the test indicates drunkeness, only 50% of them were actually drink driving.

In terms of Bayes' theorem, the relevant values are as follows:

- ▷ $pr(D) = 0.01$
- ▷ $pr(T|D) = 0.99$
- ▷ $pr(T) = 0.0198$

You can work out $pr(T)$ as just the total number of people for whom the test indicates drunkenness, divided by the total number of people in the population. Once we know the values of $pr(D)$, $pr(T|D)$, and $pr(T|\neg D)$, it's easy to work out the value of $pr(T)$. Plugging all this in to the equation gets us:

$$pr(D|T) = \frac{pr(D) \times pr(T|D)}{pr(T)}$$
$$pr(D|T) = \frac{0.01 \times 0.99}{0.0198}$$
$$pr(D|T) = 0.5$$

On the other hand, suppose that 1 in 10 drivers are drunk driving on this night. In this case, in a population of 10,000 drivers there will be 1000 drunks on the road. If every one of them were tested, the breathalyzer would correctly indicate drunkenness for 990 of them. However, of the remaining 9,000 non-drunk drivers, the test would indicate drunkenness for 90 of them. So in total, the test indicates drunkenness for 1080 drivers, of whom about 91.6% are actually drunk.

In terms of Bayes' theorem, the relevant values are as follows:

- ▷ $pr(D) = 0.1$
- ▷ $pr(T|D) = 0.99$
- ▷ $pr(T) = 0.108$

In this case,

$$pr(D|T) = \frac{pr(D) \times pr(T|D)}{pr(T)}$$
$$pr(D|T) = \frac{0.1 \times 0.99}{0.108}$$
$$pr(D|T) = 0.91666\ldots$$

The upshot here is: *don't ignore base rates*!

Chapter 7: Key Ideas

▷ **Conditional probabilities** are probabilities under a supposition, such as the probability that you are sick *given* that you have symptoms. We read '$pr(\mathcal{X}|\mathcal{Y}) = n$' as saying that the probability of $\mathcal{X}$ *given* $\mathcal{Y}$ is n.

▷ The probability of $\mathcal{X}$ conditional on $\mathcal{Y}$ is sometimes, but not always, different from the unconditional probability of $\mathcal{X}$. When $pr(\mathcal{X}|\mathcal{Y}) = pr(\mathcal{X})$, we say that $\mathcal{X}$ and $\mathcal{Y}$ are **probabilistically independent** of one another; otherwise, we say that they are **probabilistically correlated**. If $\mathcal{X}$ and $\mathcal{Y}$ are probabilistically independent of one another, then assuming $\mathcal{Y}$ makes no difference to the probability of $\mathcal{X}$.

▷ We introduced one more axiom of probability, which relates to conditional probabilities, and used it to derive several new laws of probability:

Axiom 4: Probability of a conjunction
$pr(\mathcal{X} \wedge \mathcal{Y}) = pr(\mathcal{X}|\mathcal{Y}) \times pr(\mathcal{Y})$

Law 6: Definition of conditional probability.
If $pr(\mathcal{Y}) > 0$, then $pr(\mathcal{X}|\mathcal{Y}) \ = \ pr(\mathcal{X} \wedge \mathcal{Y}) \ / \ pr(\mathcal{Y})$

Law 7: Probabilities of conjunctions with independent conjuncts.
If $\mathcal{X}$ and $\mathcal{Y}$ are independent, $pr(\mathcal{X} \wedge \mathcal{Y}) = pr(\mathcal{X}) \times pr(\mathcal{Y})$

Law 8: Bayes' Theorem
If $pr(\mathcal{Y}) > 0$, then:

$$pr(\mathcal{X}|\mathcal{Y}) = \frac{pr(\mathcal{X}) \times pr(\mathcal{Y}|\mathcal{X})}{pr(\mathcal{Y})}$$

▷ Bayes' Theorem is very useful in helping us to avoid the **base rate fallacy**. Bayes' theorem tells us that the conditional probability $pr(\mathcal{X}|\mathcal{Y})$ depends on three factors (assuming $pr(\mathcal{Y}) > 0$):

1. The unconditional probability of $\mathcal{X}$,
2. The probability of $\mathcal{Y}$ conditional on $\mathcal{X}$, and
3. The unconditional probability of $\mathcal{Y}$.

The base rate fallacy arises when we fail to consider each of these factors.

Practice Exercises

Part A
You have a standard deck of 52 cards. Of a card drawn at random, what is:

1. The probability that the card is not a jack, given that it's a face card?
2. The probability that the card is hearts or clubs, given that it's an ace?
3. The probability that the card is an ace or a king, given it's neither a 2 nor a 3?
4. The probability that the card is a black jack, given it's either red or black?
5. The probability that the card is hearts and not an ace, given that water is wet?

⋆ **Part B**
Which of the above cases involve probabilistic independence?

Part C
You know that, tonight, 1 out of 1000 drivers are drink driving. The breathalyser the police are using never fails to detect a person who is actually drunk. For 99 out of the 999 drivers who are not drunk, though, the breathalyser falsely displays drunkenness. Suppose a driver is stopped at random, and takes a breath test. The test indicates that he or she is drunk. We assume you don't know anything else about the driver. What is the probability that he or she really is drunk?

⋆ **Part D**
Your friend is worried she might have caught the flu. This season, about 1 in every 100 people will catch it. As yet, she hasn't shown any particular symptoms, nor any particular reason to think that she's more likely to have caught it than anyone else. Nevertheless, she has a test which boasts high levels of accuracy. If you have the flu, the test will return a positive result 99% of the time. If you don't have the flu, it will return a negative result 90% of the time.

Letting F stand for 'She has the flu', and T stand for 'The test returns positive', what are the values of the following (rounded to the 4th decimal):

1. $pr(F)$
2. $pr(\neg F)$
3. $pr(T|F)$
4. $pr(\neg T|\neg F)$
5. $pr(T|\neg F)$
6. $pr(T)$
7. $pr(\neg T)$
8. $pr(F|T)$
9. $pr(\neg F|\neg T)$

Part E
Explain how one might try to derive the following additional laws of probability, which hold for any pair of propositions $\mathcal{X}$ and $\mathcal{Y}$. You can use any of the axioms and laws of probability already given in this chapter and the last chapter.

1. $pr(\mathcal{X}|\mathcal{Y}) \times pr(\mathcal{Y}) = pr(\mathcal{Y}|\mathcal{X}) \times pr(\mathcal{X})$
2. $pr(\mathcal{X}) = (pr(\mathcal{X}|\mathcal{Y}) \times pr(\mathcal{Y})) + (pr(\mathcal{X}|\neg\mathcal{Y}) \times pr(\neg\mathcal{Y}))$

Part III

Appendix

Appendix A

Proofs in Propositional Logic

In this Appendix, we will look at the standard **proof theory** that's applicable when we're using $\mathcal{L}_S$. You will not be tested on anything in this chapter, but it's helpful stuff to know—especially if you want to continue your studies in logic!

Consider two arguments in $\mathcal{L}_S$:

Argument **A**

P1 $(P \vee Q)$
P2 $\neg P$

C Q

Argument **B**

P1 $(P \rightarrow Q)$
P2 P

C Q

Clearly, these are valid arguments. You can confirm that they are valid by constructing four-line truth tables. Argument **A** makes use of an inference form that is always valid: given a disjunction and the negation of one of the disjuncts, the other disjunct follows as a valid consequence. This rule is called **disjunctive syllogism**.

Argument **B** makes use of a different valid form: given a conditional and its antecedent, the consequent follows as a valid consequence. This is called **modus ponens**.

When we construct truth tables, we do not need to give names to different inference forms. There is no reason to distinguish modus ponens from a disjunctive syllogism. For this same reason, however, the method of truth tables does not clearly show *why* an argument is valid. If you were to do a 1024-line truth table for an argument that contains ten sentence letters, then you could check to see if there were any lines on which the premises were all true and the conclusion were false. If you did not see such a line and provided you made no mistakes in constructing the table, then you would know that the argument was valid. Yet you would not be able to say anything further about why this particular argument was a valid argument form.

The aim of a **proof system** is to show that particular arguments are valid in a way that allows us to understand the reasoning involved in the argument. We begin with basic argument forms, like disjunctive syllogism and modus ponens. These forms can then be combined to make more complicated arguments, like this one:

P1 $(\neg L \to (J \vee L))$
P2 $\neg L$

C J

By modus ponens, P1 and P2 entail $(J \vee L)$. This is an **intermediate conclusion**. It follows logically from the premises, but it is not the conclusion we want. Now $(J \vee L)$ and P2 entail the conclusion, J, by disjunctive syllogism. We do not need a new rule for this argument. The proof of the argument shows that it is really just a combination of rules we have already introduced.

Formally, a **proof** is a sequence of sentences. The first sentences of the sequence are often called **assumptions**; these are the premises of a valid argument. Every sentence later in the sequence follows from earlier sentences by one of the rules of proof. The final sentence of the sequence is the conclusion of a valid argument.

Very generally, a **rule of proof** is a general rule that says: if you have already written down a line or lines with such-and-such a form in your proof, then you are permitted to write down a new line with such-and-such a form as well. Some rules of proof take you from a single line in a proof to a new line, while others require two or more lines before they can be applied. Every rule of proof outlined in this chapter corresponds to a valid inference. The idea is that we should eventually be able to chain together sequences of valid inferences using these rules of proof to be able to prove the validity of any valid argument.

This chapter develops a proof system for $\mathcal{L}_S$, which is then extended to cover $\mathcal{L}_P$ (and $\mathcal{L}_P$ with identity) in Chapter 9.

1. Basic rules for $\mathcal{L}_S$

In designing a proof system, we could just start with disjunctive syllogism and modus ponens. Whenever we discovered a valid argument which could not be proven with rules we already had, we could introduce new rules. Proceeding in this way, we would have an unsystematic grab bag of rules. We might accidentally add some strange rules, and we would surely end up with more rules than we need.

Instead, we will develop what is called a **natural deduction** system. In a natural deduction system, there will be two rules for each logical operator: an **introduction** rule that allows us to prove a sentence that has it as the main logical operator; and an **elimination** rule that allows us to prove something given a sentence that has it as the main logical operator.

In addition to the rules for each logical operator, we will also have a reiteration rule. If you already have shown something in the course of a proof, the reiteration rule allows you to repeat it on a new line. For instance:

1	$\mathcal{X}$	
2	$\mathcal{X}$	R 1

This says: if you have any sentence $\mathcal{X}$ written down as a line in your proof, you're allowed to re-write it again on a separate (later) line. When we add a line to a proof, we write the rule that justifies that line. We also write the numbers of the lines to which the rule was applied. The reiteration rule above is justified by one line, the line that you are reiterating.

So the 'R 1' on line 2 of the proof means that the line is justified by the reiteration rule (which we'll label 'R'), applied to line 1.

Obviously, the reiteration rule will never allow us to prove anything *new*. For that, we will need more rules. The remainder of this chapter will give introduction and elimination rules for each of the sentential connectives introduced in Chapter 2 and Chapter 3. This will give us a complete proof system for $\mathcal{L}_S$.

Conjunction

Think for a moment: what would you need to show in order to prove $(E \wedge F)$?

Of course, you could show $(E \wedge F)$ by proving E and separately proving F. This holds even if the two conjuncts are not atomic sentences. For instance, if you can prove $((A \vee J) \rightarrow V)$ and $((V \rightarrow L) \leftrightarrow (F \vee N))$, then you have effectively proven the following as well:

$$(((A \vee J) \rightarrow V) \wedge ((V \rightarrow L) \leftrightarrow (F \vee N)))$$

So this will be our conjunction introduction rule, which we abbreviate $\wedge$I:

$$\begin{array}{r|ll} m & \mathcal{X} & \\ n & \mathcal{Y} & \\ & (\mathcal{X} \wedge \mathcal{Y}) & \wedge\text{I } m, n \end{array}$$

This says: if you have two lines in your proof, one of which says $\mathcal{X}$ and the other of which says $\mathcal{Y}$, then you're allowed to write down on a new line of proof $(\mathcal{X} \wedge \mathcal{Y})$. A line of proof must be justified by some rule, and for the new line we justify it using $\wedge$I m, n. This means: "the rule of conjunction introduction applied to lines m and n". Here, we're using m and n as variables, not real line numbers; m is some line and n is some other line. In an actual proof, the lines are numbered $1, 2, 3, \ldots$ and rules must be applied to specific line numbers. When we define the rule, however, we use variables to underscore the point that the rule may be applied to any two lines that are already in the proof. If you have K on line 8 and L on line 15, you can prove $(K \wedge L)$ at some later point in the proof with the justification $\wedge$I 8, 15.

Now, consider the elimination rule for conjunction. What are you entitled to conclude from a sentence like $(E \wedge F)$? Surely, you are entitled to conclude E; if $(E \wedge F)$ were true, then E would be true. Similarly, you are entitled to conclude F. This will be our conjunction elimination rule, which we abbreviate $\wedge$E:

$$\begin{array}{r|ll} m & (\mathcal{X} \wedge \mathcal{Y}) & \\ & \mathcal{X} & \wedge\text{E } m \\ & \mathcal{Y} & \wedge\text{E } m \end{array}$$

This says: if you have a line in your proof of the form $(\mathcal{X} \wedge \mathcal{Y})$, then you're allowed to write down either $\mathcal{X}$ or $\mathcal{Y}$ on a new line of proof below it. That is, when you have a conjunction on some line of a proof, you can use $\wedge$E to derive either of the conjuncts. The $\wedge$E rule requires only one line of proof to be applied, so we only ever write one line number as the justification.

Even with just these two rules, we can provide some proofs. Consider:

P1 $((A \vee B) \to (C \vee D)) \wedge ((E \vee F) \to (G \vee H))$

C $((E \vee F) \to (G \vee H)) \wedge ((A \vee B) \to (C \vee D))$

The main logical operator in both the premise and conclusion is conjunction. Since conjunction is symmetric, the argument is obviously valid. In order to provide a proof, we begin by writing down the premises—in this case, there's only one. After the premises, we draw a horizontal line—everything below this line must be justified by a rule of proof. So the beginning of the proof is:

1	$((A \vee B) \to (C \vee D)) \wedge ((E \vee F) \to (G \vee H))$	

From the premise, we can get each of the conjuncts by $\wedge$E:

1	$((A \vee B) \to (C \vee D)) \wedge ((E \vee F) \to (G \vee H))$	
2	$((A \vee B) \to (C \vee D))$	$\wedge$E 1
3	$((E \vee F) \to (G \vee H))$	$\wedge$E 1

The rule $\wedge$I requires that we have each of the conjuncts available somewhere in the proof. They can be separated from one another, and they can appear in any order. So by applying the $\wedge$I rule to lines 3 and 2, we arrive at the desired conclusion. The finished proof looks like this:

1	$((A \vee B) \to (C \vee D)) \wedge ((E \vee F) \to (G \vee H))$	
2	$((A \vee B) \to (C \vee D))$	$\wedge$E 1
3	$((E \vee F) \to (G \vee H))$	$\wedge$E 1
4	$((E \vee F) \to (G \vee H)) \wedge ((A \vee B) \to (C \vee D))$	$\wedge$I 3, 2

This proof is trivial, but it shows how we can use rules of proof together to demonstrate the validity of an argument form. Also: using a truth table to show that this argument is valid would have required a staggering 256 lines, since there are eight sentence letters in the argument.

Disjunction

If M were true, then $(M \vee N)$ would also be true. This can be easily checked by $\vee$'s truth table: a disjunction is true whenever at least one of its disjuncts is true. So the disjunction introduction rule ($\vee$I) allows us to derive a disjunction if we have assumed the truth of either one of the two disjuncts:

m	$\mathcal{X}$	
	$(\mathcal{X} \vee \mathcal{Y})$	$\vee$I m
	$(\mathcal{Y} \vee \mathcal{X})$	$\vee$I m

This says: if you have a sentence $\mathcal{X}$ written down, you are allowed to write on a new line another sentence of either the form $(\mathcal{X} \vee \mathcal{Y})$, or $(\mathcal{Y} \vee \mathcal{X})$. In this case, $\mathcal{Y}$ can be any sentence whatsoever. So the following is a legitimate proof:

$$
\begin{array}{r|ll}
1 & M & \\
\hline
2 & (M \vee (A \wedge B) & \vee\text{I } 1
\end{array}
$$

It may seem odd that just by knowing M we can derive a conclusion that includes things like A, B, and the rest—sentences that intuitively have nothing to do with M. Yet the conclusion follows immediately by $\vee$I. This is as it should be: the truth conditions for the disjunction mean that, if $\mathcal{X}$ is true, then $(\mathcal{X} \vee \mathcal{Y})$ is true regardless of what $\mathcal{Y}$ is. So the conclusion could not be false if the premise were true; the argument is valid.

Now consider the disjunction elimination rule. What can you conclude from $(M \vee N)$? You cannot conclude M. It might be M truth that makes $(M \vee N)$ true, as in the example above, but it might not. From $(M \vee N)$ alone, you cannot conclude anything about either M or N specifically. If you also knew that N was false, however, then you would be able to conclude M.

This is just disjunctive syllogism, it will be the disjunction elimination rule ($\vee$E). Strictly speaking, there are two rules here, depending on what disjunct is being negated:

$$
\begin{array}{r|ll}
m & (\mathcal{X} \vee \mathcal{Y}) & \\
n & \neg\mathcal{Y} & \\
 & \mathcal{X} & \vee\text{E } m,\, n
\end{array}
\qquad
\begin{array}{r|ll}
m & (\mathcal{X} \vee \mathcal{Y}) & \\
n & \neg\mathcal{X} & \\
 & \mathcal{Y} & \vee\text{E } m,\, n
\end{array}
$$

Each rule says: if you have a line of proof of the form $(\mathcal{X} \vee \mathcal{Y})$, and another line of proof of the form $\neg\mathcal{X}$ (or $\neg\mathcal{Y}$), you're allowed to write $\mathcal{Y}$ or $\mathcal{X}$ down on a new line.

Conditional

Consider this argument:

$$
\begin{array}{rl}
\textbf{P1} & (R \vee F) \\
\hline
\textbf{C} & (\neg R \to F)
\end{array}
$$

The argument is certainly a valid one. What should the conditional introduction rule be, such that we can draw this conclusion?

We begin the proof by writing down the premise of the argument and drawing a horizontal line, like this:

$$
\begin{array}{r|l}
1 & (R \vee F) \\
\hline
\end{array}
$$

If we had $\neg R$ as a further premise, we could derive F by the $\vee$E rule. We do not have $\neg R$ as a premise of this argument, nor can we derive it directly from the premise we do have—so we cannot simply prove F. What we will do instead is start a **subproof**, a proof within the main proof. When we start a subproof, we draw another vertical line to indicate that we are no longer in the main proof. Then we write in an assumption for the subproof. This can be anything we want. Here, it will be helpful to assume $\neg R$. (Over time, you'll get the hang of what you should assume for subproofs.) Our proof now looks like this:

1	$(R \vee F)$	
2	$\quad \neg R$	

It is important to notice that we are not claiming to have proven $\neg R$. We do not need to write in any justification for the assumption line of a subproof. You can think of the subproof as posing the question: what could we show *if* we assumed that $\neg R$ were true? For one thing, we can derive F. So we do:

1	$(R \vee F)$	
2	$\quad \neg R$	
3	$\quad F$	$\vee$E 1, 2

This has shown that *if* we had $\neg R$ as a premise, *then* we could prove F. In effect, we have proven $(\neg R \to F)$. So the conditional introduction rule ($\to$I) will allow us to close the subproof and derive $(\neg R \to F)$ in the main proof. Our final proof looks like this:

1	$(R \vee F)$	
2	$\quad \neg R$	
3	$\quad F$	$\vee$E 1, 2
4	$(\neg \mathrm{R} \to \mathrm{F})$	$\to$I 2–3

Notice that the justification for applying the $\to$I rule is the entire subproof. Usually that will be more than just two lines.

It may seem as if the ability to assume anything at all in a subproof would lead to chaos: Does it allow you to prove any conclusion from any premises? The answer is no, it does not. Consider this proof:

1	$\mathcal{X}$	
2	$\quad \mathcal{Y}$	
3	$\quad \mathcal{Y}$	R 2

It may seem as if this is a proof that you can derive any conclusions $\mathcal{Y}$ from any premise $\mathcal{X}$. When the vertical line for the subproof ends, the subproof is *closed.* In order to complete a proof, you must close all of the subproofs. And you cannot close the subproof and use the R rule again on line 4 to derive $\mathcal{Y}$ in the main proof. Once you close a subproof, you cannot refer back to individual lines inside it.

Closing a subproof is called *discharging* the assumptions of that subproof. So we can put the point this way: You cannot complete a proof until you have discharged all of the assumptions besides the original premises of the argument.

Of course, it is legitimate to do this:

1	$\mathcal{X}$	
2	$\quad \mathcal{Y}$	
3	$\quad \mathcal{Y}$	R 2
4	$(\mathcal{Y} \to \mathcal{Y})$	$\to$I 2–3

This should not seem so strange, though. Since $(\mathcal{Y} \to \mathcal{Y})$ is a tautology, no particular premises should be required to validly derive it. (Indeed, as we will see, a tautology follows from any premises.)

Put in a general form, the $\to$I rule looks like this:

m	$\quad \mathcal{X}$	
n	$\quad \mathcal{Y}$	
	$(\mathcal{X} \to \mathcal{Y})$	$\to$I m–n

This says that if you are able to derive a sentence $\mathcal{Y}$ from a subproof assumption $\mathcal{X}$, then you are allowed to write down on a later line of proof $(\mathcal{X} \to \mathcal{Y})$. When we introduce a subproof, we typically write what we want to derive in the column. This is just so that we do not forget why we started the subproof if it goes on for five or ten lines.

Although it is always permissible to open a subproof with any assumption you please, there is some strategy involved in picking a useful assumption. Starting a subproof with an arbitrary, wacky assumption would just waste lines of the proof. In order to derive a conditional by the $\to$I, for instance, you must assume the antecedent of the conditional in a subproof. In general, if you want to prove a conditional, it's usually a good strategy to start by assuming the antecedent of that conditional.

The $\to$I rule also requires that the consequent of the conditional be the last line of the subproof. It is always permissible to close a subproof and discharge its assumptions, but it will not be helpful to do so until you get what you want.

Now consider the conditional elimination rule. Nothing follows from $(M \to N)$ alone, but if we have both $(M \to N)$ *and* M, then we can conclude N. This rule, *modus ponens*, will be the conditional elimination rule ($\to$E).

m	$(\mathcal{X} \to \mathcal{Y})$	
n	$\mathcal{X}$	
	$\mathcal{Y}$	$\to$E m, n

That is, if you have two lines of proof, one of the form $(\mathcal{X} \to \mathcal{Y})$ and the other of the form $\mathcal{X}$ (i.e., the same as the antecedent in $(\mathcal{X} \to \mathcal{Y})$), then you're allowed to write down $\mathcal{Y}$ in a later line of proof.

Now that we have rules for the conditional, consider this argument:

P1 $(P \to Q)$
P2 $(Q \to R)$

C $(P \to R)$

We begin the proof by writing the two premises as assumptions. Since the main logical operator in the conclusion is a conditional, we can expect to use the →I rule. For that, we need a subproof—so we write in the antecedent of the conditional as assumption of a subproof:

$$
\begin{array}{r|l}
1 & (P \to Q) \\
2 & (Q \to R) \\
\hline
3 & \quad \underline{\vert\, P} \\
\end{array}
$$

We made P available by assuming it in a subproof, allowing us to use →E on the first premise. This gives us Q, which allows us to use →E on the second premise. Having derived R, we close the subproof. By assuming P we were able to prove R, so we apply the →I rule and finish the proof.

$$
\begin{array}{r|ll}
1 & (P \to Q) & \\
2 & (Q \to R) & \\
\hline
3 & \quad \vert\, \underline{P} & \text{want } R \\
4 & \quad \vert\, Q & \to\text{E } 1, 3 \\
5 & \quad \vert\, R & \to\text{E } 2, 4 \\
6 & (P \to R) & \to\text{I } 3\text{–}5 \\
\end{array}
$$

Biconditional

The rules for the biconditional will be like double-barreled versions of the rules for the conditional.

In order to derive $(W \leftrightarrow X)$, for instance, you must be able to prove X by assuming W *and* prove W by assuming X. The biconditional introduction rule (↔I) requires two subproofs. The subproofs can come in any order, and the second subproof does not need to come immediately after the first—but schematically, the rule works like this:

$$
\begin{array}{r|ll}
m & \quad \vert\, \underline{\mathcal{X}} & \text{want } \mathcal{Y} \\
n & \quad \vert\, \mathcal{Y} & \\
p & \quad \vert\, \underline{\mathcal{Y}} & \text{want } \mathcal{X} \\
q & \quad \vert\, \mathcal{X} & \\
 & (\mathcal{X} \leftrightarrow \mathcal{Y}) & \leftrightarrow\text{I } m\text{–}n,\ p\text{–}q \\
\end{array}
$$

The biconditional elimination rule (↔E) lets you do a bit more than the conditional rule. If you have the left-hand side of the biconditional, you can derive the right-hand side. If you have the right-hand side, you can derive the left-hand side. This is the rule:

$$
\begin{array}{r|ll}
m & (\mathcal{X} \leftrightarrow \mathcal{Y}) & \\
n & \mathcal{X} & \\
 & \mathcal{Y} & \leftrightarrow\text{E } m, n \\
\end{array}
\qquad\qquad
\begin{array}{r|ll}
m & (\mathcal{X} \leftrightarrow \mathcal{Y}) & \\
n & \mathcal{Y} & \\
 & \mathcal{X} & \leftrightarrow\text{E } m, n \\
\end{array}
$$

Negation

Here is a simple mathematical argument in English:

P1 Assume there is some greatest natural number. Call it α.
P2 That number plus one is also a natural number.
P3 Obviously, $a + 1 > a$.
P4 So there is a natural number greater than α.
P5 This is impossible, as α is assumed to be the greatest natural number.

C There is no greatest natural number.

This argument form is traditionally called a *reductio*. Its full Latin name is *reductio ad absurdum*, which means 'reduction to absurdity.' In a reductio, we assume something for the sake of argument—for example, that there is a greatest natural number. Then we show that the assumption leads to two contradictory sentences—for example, that α is the greatest natural number and that it is not. In this way, we show that the original assumption must have been false.

The basic rules for negation will allow for arguments like this. If we assume something and show that it leads to contradictory sentences, then we have proven the negation of the assumption. This is the negation introduction (¬I) rule:

m	$\quad \mathcal{X}$	for reductio
n	$\quad \mathcal{Y}$	
$n+1$	$\quad \neg\mathcal{Y}$	
$n+2$	$\neg\mathcal{X}$	¬I m–$n+1$

That is, if you can derive a contradiction ($\mathcal{Y}$ and $\neg\mathcal{Y}$) from a subproof assumption $\mathcal{X}$, then you can write down on a later line of proof $\neg\mathcal{X}$. For the rule to apply, the last two lines of the subproof must be an **explicit contradiction**: some sentence followed on the next line by its negation. We write 'for reductio' as a note to ourselves, a reminder of why we started the subproof. It is not formally part of the proof, and you can leave it out if you find it distracting.

To see how the rule works, suppose we want to prove the law of non-contradiction: $\neg(G \wedge \neg G)$. We can prove this without any premises by immediately starting a subproof. We want to apply ¬I to the subproof, so we assume $(G \wedge \neg G)$. We then get an explicit contradiction by ∧E. The proof looks like this:

1	$\quad (G \wedge \neg G)$	for reductio
2	$\quad G$	∧E 1
3	$\quad \neg G$	∧E 1
4	$\neg(G \wedge \neg G)$	¬I 1–3

The ¬E rule will work in much the same way. If we assume $\neg\mathcal{X}$ and show that it leads to a contradiction, we have effectively proven $\mathcal{X}$. So the rule looks like this:

m	$\quad\mid\quad\mid \neg\mathcal{X}$	for reductio
n	$\quad\mid\quad\mid \mathcal{Y}$	
$n+1$	$\quad\mid\quad\mid \neg\mathcal{Y}$	
$n+2$	$\quad\mid \mathcal{X}$	$\neg$E m–$n+1$

2. Derived rules

The rules of the natural deduction system are meant to be systematic. There is an introduction and an elimination rule for each logical operator, but why these basic rules rather than some others? Many natural deduction systems have a disjunction elimination rule that works like this:

m	$\mid (\mathcal{X} \vee \mathcal{Y})$	
n	$\mid (\mathcal{X} \to \mathcal{C})$	
o	$\mid (\mathcal{Y} \to \mathcal{C})$	
	$\mid \mathcal{C}$	DIL m, n, o

Let's call this rule *Dilemma* (DIL) It might seem as if there will be some proofs that we cannot do with our proof system, because we do not have this as a basic rule. Yet this is not the case. Any proof that you can do using the Dilemma rule can be done with basic rules of our natural deduction system. Consider this proof:

1	$\mid (\mathcal{X} \vee \mathcal{Y})$	
2	$\mid (\mathcal{X} \to \mathcal{C})$	
3	$\mid (\mathcal{Y} \to \mathcal{C})$	want $\mathcal{C}$
4	$\mid\quad\mid \neg\mathcal{C}$	for reductio
5	$\mid\quad\mid\quad\mid \mathcal{X}$	for reductio
6	$\mid\quad\mid\quad\mid \mathcal{C}$	$\to$E 2, 5
7	$\mid\quad\mid\quad\mid \neg\mathcal{C}$	R 4
8	$\mid\quad\mid \neg\mathcal{X}$	$\neg$I 5–7
9	$\mid\quad\mid\quad\mid \mathcal{Y}$	for reductio
10	$\mid\quad\mid\quad\mid \mathcal{C}$	$\to$E 3, 9
11	$\mid\quad\mid\quad\mid \neg\mathcal{C}$	R 4
12	$\mid\quad\mid \mathcal{Y}$	$\vee$E 1, 8
13	$\mid\quad\mid \neg\mathcal{Y}$	$\neg$I 9–11
14	$\mid \mathcal{C}$	$\neg$E 4–13

$\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{C}$ are meta-variables; they are not symbols of $\mathcal{L}_S$, but stand-ins for arbitrary sentences symbolised in $\mathcal{L}_S$. So this is not, strictly speaking, a proof in $\mathcal{L}_S$. It is more like a recipe. It provides a pattern that can prove anything that the Dilemma rule can prove, using only the basic rules of $\mathcal{L}_S$. This means that the Dilemma rule is not really necessary. Adding it to the list of basic rules would not allow us to derive anything that we could not derive without it.

Nevertheless, the Dilemma rule would be convenient. It would allow us to do in one line what requires eleven lines and several nested subproofs with the basic rules. So we will add it to the proof system as a **derived rule**. A derived rule is a rule of proof that does not make any new proofs possible. Anything that can be proven with a derived rule can be proven without it. You can think of a short proof using a derived rule as shorthand for a longer proof that uses only the basic rules. Any time you use the Dilemma rule, you could always take ten extra lines and prove the same thing without it.

For the sake of convenience, we will add several other derived rules. One is *modus tollens* (MT).

m	$(\mathcal{X} \rightarrow \mathcal{Y})$	
n	$\neg\mathcal{Y}$	
	$\neg\mathcal{X}$	MT m, n

We leave the proof of this rule as an exercise. Note that if we had already proven the MT rule, then the proof of the DIL rule could have been done in only five lines.

We also add *hypothetical syllogism* (HS) as a derived rule. We have already given a proof of it on p. 108.

m	$(\mathcal{X} \rightarrow \mathcal{Y})$	
n	$(\mathcal{Y} \rightarrow \mathcal{C})$	
	$(\mathcal{X} \rightarrow \mathcal{C})$	HS m, n

3. Rules of replacement

Consider how you would prove this argument:

P1 $(F \rightarrow (G \wedge H))$

C $(F \rightarrow G)$

Perhaps it is tempting to write down the premise and apply the $\wedge$E rule to the conjunction $(G \wedge H)$. This is impermissible, however, because the basic rules of proof can only be applied to whole sentences. We need to get $(G \wedge H)$ on a line by itself. We can prove the argument in this way:

1	$(F \rightarrow (G \wedge H))$	
2	$\quad F$	want G
3	$\quad (G \wedge H)$	$\rightarrow$E 1, 2
4	$\quad G$	$\wedge$E 3
5	$(F \rightarrow G)$	$\rightarrow$I 2–4

We will now introduce some derived rules that may be applied to a *part* of a sentence. These are called **rules of replacement**, because they can be used to replace part of a sentence with a logically equivalent expression. One simple rule of replacement is *commutivity* (abbreviated Comm), which says that we can swap the order of conjuncts in a conjunction or the order of disjuncts in a disjunction. We define the rule this way:

$$\begin{aligned} (\mathcal{X} \wedge \mathcal{Y}) &\Longleftrightarrow (\mathcal{Y} \wedge \mathcal{X}) \\ (\mathcal{X} \vee \mathcal{Y}) &\Longleftrightarrow (\mathcal{Y} \vee \mathcal{X}) \\ (\mathcal{X} \leftrightarrow \mathcal{Y}) &\Longleftrightarrow (\mathcal{Y} \leftrightarrow \mathcal{X}) \quad \text{Comm} \end{aligned}$$

The bold arrow means that you can take a subformula on one side of the arrow and replace it with the subformula on the other side. The arrow is double-headed because rules of replacement work in both directions.

Consider this argument:

P1 $((M \vee P) \to (P \wedge M))$

C $((P \vee M) \to (M \wedge P))$

It is possible to give a proof of this using only the basic rules, but it will be long and inconvenient. With the Comm rule, we can provide a proof easily:

1	$((M \vee P) \to (P \wedge M))$	
2	$((P \vee M) \to (P \wedge M))$	Comm 1
3	$((P \vee M) \to (M \wedge P))$	Comm 2

Another useful rule of replacement is *double negation* (DN). With the DN rule, you can remove or insert a pair of negations anywhere in a symbolized sentence. This is the rule:

$$\neg\neg\mathcal{X} \Longleftrightarrow \mathcal{X} \quad \text{DN}$$

Two more replacement rules are called *De Morgan's Laws*, named for the 19th-century British logician August De Morgan. (Although De Morgan did discover these laws, he was not the first to do so.) The rules capture useful relations between negation, conjunction, and disjunction. Here are the rules, which we abbreviate DeM:

$$\begin{aligned} \neg(\mathcal{X} \vee \mathcal{Y}) &\Longleftrightarrow (\neg\mathcal{X} \wedge \neg\mathcal{Y}) \\ \neg(\mathcal{X} \wedge \mathcal{Y}) &\Longleftrightarrow (\neg\mathcal{X} \vee \neg\mathcal{Y}) \quad \text{DeM} \end{aligned}$$

Because $(\mathcal{X} \to \mathcal{Y})$ is a *material conditional*, it is equivalent to $(\neg\mathcal{X} \vee \mathcal{Y})$. A further replacement rule captures this equivalence. We abbreviate the rule MC, for 'material conditional.' It takes two forms:

$$\begin{aligned} (\mathcal{X} \to \mathcal{Y}) &\Longleftrightarrow (\neg\mathcal{X} \vee \mathcal{Y}) \\ (\mathcal{X} \vee \mathcal{Y}) &\Longleftrightarrow (\neg\mathcal{X} \to \mathcal{Y}) \quad \text{MC} \end{aligned}$$

Now consider this argument:

P1	$\neg(P \to Q)$
C	$(P \land \neg Q)$

As always, we could prove this argument using only the basic rules. With rules of replacement, though, the proof is much simpler:

1	$\neg(P \to Q)$	
2	$\neg(\neg P \lor Q)$	MC 1
3	$(\neg\neg P \land \neg Q)$	DeM 2
4	$(P \land \neg Q)$	DN 3

A final replacement rule captures the relation between conditionals and biconditionals. We will call this rule *biconditional exchange* and abbreviate it $\leftrightarrow$ex.

$$((\mathcal{X} \to \mathcal{Y}) \land (\mathcal{Y} \to \mathcal{X})) \iff (\mathcal{X} \leftrightarrow \mathcal{Y}) \quad \leftrightarrow\text{ex}$$

4. Proof strategy

There is no simple recipe for proofs, and there is no substitute for practice. Here, though, are some rules of thumb and strategies to keep in mind.

Work backwards from what you want. The ultimate goal is to derive the conclusion. Look at the conclusion and ask what the introduction rule is for its main logical operator. This gives you an idea of what should happen *just before* the last line of the proof. Then you can treat this line as if it were your goal. Ask what you could do to derive this new goal.

For example: If your conclusion is a conditional $(\mathcal{X} \to \mathcal{Y})$, plan to use the $\to$I rule. This requires starting a subproof in which you assume $\mathcal{X}$. In the subproof, you want to derive $\mathcal{Y}$.

Work forwards from what you have. When you are starting a proof, look at the premises; later, look at the things that you have derived so far. Think about the elimination rules for the main logical connectives. These will tell you what your options are.

For a short proof, you might be able to eliminate the premises and introduce the conclusion. A long proof is formally just a number of short proofs linked together, so you can fill the gap by alternately working back from the conclusion and forward from the premises.

Change what you are looking at. Replacement rules can often make your life easier. If a proof seems impossible, try out some different substitutions. For example: It is often difficult to prove a disjunction using the basic rules. If you want to show $(\mathcal{X} \lor \mathcal{Y})$, it is often easier to show $(\neg\mathcal{X} \to \mathcal{Y})$ and use the MC rule.

Some replacement rules should become second nature. If you see a negated disjunction, for instance, you should immediately think of DeMorgan's rule.

Do not forget indirect proof. If you cannot find a way to show something directly, try assuming its negation. Remember that most proofs can be done either indirectly or directly. One way might be easier—or perhaps one sparks your imagination more than the other—but either one is formally legitimate.

Repeat as necessary. Once you have decided how you might be able to get to the conclusion, ask what you might be able to do with the premises. Then consider the target sentences again and ask how you might reach them.

Persist. Try different things. If one approach fails, then try something else.

Practice Exercises

⋆ Part A

Provide a justification (rule and line numbers) for each line of proof that requires one.

1	$(W \to \neg B)$
2	$(A \wedge W)$
3	$(B \vee (J \wedge K))$
4	W
5	$\neg B$
6	$(J \wedge K)$
7	K

1	$(L \leftrightarrow \neg O)$
2	$(L \vee \neg O)$
3	$\quad \neg L$
4	$\quad \neg O$
5	$\quad L$
6	$\quad \neg L$
7	L

1	$(Z \to (C \wedge \neg N))$
2	$(\neg Z \to (N \wedge \neg C))$
3	$\quad \neg(N \vee C)$
4	$\quad (\neg N \wedge \neg C)$
5	$\quad\quad Z$
6	$\quad\quad (C \wedge \neg N)$
7	$\quad\quad C$
8	$\quad\quad \neg C$
9	$\quad \neg Z$
10	$\quad (N \wedge \neg C)$
11	$\quad N$
12	$\quad \neg N$
13	$(N \vee C)$

⋆ Part B

Give a proof for each argument in $\mathcal{L}_S$.

1. $(K \wedge L)$
   therefore $(K \leftrightarrow L)$

2. $(A \to (B \to C))$
   therefore $((A \wedge B) \to C)$

3. $(P \wedge (Q \vee R))$
   $(P \to \neg R)$
   therefore $(Q \vee E)$

4. $((C \wedge D) \vee E)$
   therefore $(E \vee D)$

5. $(\neg F \to G)$
   $(F \to H)$
   therefore $(G \vee H)$

6. $((X \wedge Y) \vee (X \wedge Z))$
   $\neg(X \wedge D)$
   $(D \vee M)$ therefore M

Part C

Give a proof for each argument in $\mathcal{L}_S$.

1. $(Q \rightarrow (Q \wedge \neg Q))$
   therefore $\neg Q$

2. $(J \rightarrow \neg J)$
   therefore $\neg J$

3. $(E \vee F)$
   $(F \vee G)$
   $\neg F$
   therefore $(E \wedge G)$

4. $(A \leftrightarrow B)$
   $(B \leftrightarrow C)$
   therefore $(A \leftrightarrow C)$

5. $(M \vee (N \rightarrow M))$
   therefore $(\neg M \rightarrow \neg N)$

6. $(S \leftrightarrow T)$
   therefore $(S \leftrightarrow (T \vee S))$

7. $((M \vee N) \wedge (O \vee P))$, $(N \rightarrow P)$
   $\neg P$, therefore $(M \wedge O)$

8. $((Z \wedge K) \vee (K \wedge M))$, $(K \rightarrow D)$
   therefore D

Part D

Show that each of the following pairs are provably equivalent in $\mathcal{L}_S$.

1. $\neg\neg\neg\neg G$
   G

2. $(T \rightarrow S)$
   $(\neg S \rightarrow \neg T)$

3. $(R \leftrightarrow E)$
   $(E \leftrightarrow R)$

4. $(\neg G \leftrightarrow H)$
   $\neg(G \leftrightarrow H)$

5. $(U \rightarrow I)$
   $\neg(U \wedge \neg I)$

Part E

Provide proofs to show each of the following.

1. $(M \wedge (\neg N \rightarrow \neg M))$ implies $((N \wedge M) \vee \neg M)$
2. $(C \rightarrow (E \wedge G))$ and $(\neg C \rightarrow G)$ imply G
3. $((Z \wedge K) \leftrightarrow (Y \wedge M))$ and $(D \wedge (D \rightarrow M))$ imply $(Y \rightarrow Z)$
4. $((W \vee X) \vee (Y \vee Z))$, $(X \rightarrow Y)$ and $\neg Z$ imply $(W \vee Y)$

Part F
For the following, provide proofs using only the basic rules. The proofs will be longer than proofs of the same claims would be using the derived rules.

1. Show that MT is a legitimate derived rule. Using only the basic rules, prove the following: $(A \rightarrow B)$, $\neg B$, therefore $\neg A$.
2. Show that Comm is a legitimate rule for the biconditional. Using only the basic rules, prove that $(A \leftrightarrow\ )$ and $(B \leftrightarrow A)$ are equivalent.
3. Using only the basic rules, prove the following instance of DeMorgan's Laws: $(\neg A \wedge \neg B)$, therefore $\neg(A \vee B)$
4. Show that $\leftrightarrow$ex is a legitimate derived rule. Using only the basic rules, prove that $(D \leftrightarrow E)$ and $((D \rightarrow E) \wedge (E \rightarrow D))$ are equivalent.

Appendix B

Proofs in Quantificational Logic

In this supplementary chapter, we will develop the classical proof theory that's applicable when we're dealing with sentences formalized in $\mathcal{L}_P$.

For proofs in $\mathcal{L}_P$, we use all of the basic rules of $\mathcal{L}_S$ plus four new basic rules: these will be the introduction and elimination rules for each of the quantifiers. Since all of the derived rules of $\mathcal{L}_S$ are derived from the basic rules, they will also hold in $\mathcal{L}_P$. We will add another derived rule, a replacement rule called *quantifier negation.*

1. Substitution instances

In order to concisely state the rules for the quantifiers, we need a way to mark the relation between quantified sentences and their instances. For example, the sentence 'Pa' is a particular instance of the general claim '$\forall x(Px)$.'

For a well-formed formula $\mathcal{A}$ written in $\mathcal{L}_P$, a constant 'c,' and a variable 'x,' define '$\mathcal{A}\boxed{\chi\Rightarrow c}$' to mean the formula that we get by replacing every occurrence of 'x' in '$\mathcal{A}$' with 'c.' '$\mathcal{A}\boxed{\chi\Rightarrow c}$' is called a **substitution instance** of '$\forall x\mathcal{A}$' and '$\exists x\mathcal{A}$,' and 'c' is called the **instantiating constant**. For example:

- ▷ '$Aa \to Ba$,' '$Af \to Bf$,' and '$Ak \to Bk$' are all substitution instances of '$\forall x(Ax \to Bx)$'; the instantiating constants are 'a,' 'f,' and 'k,' respectively.
- ▷ 'Raj,' 'Rdj,' and 'Rjj' are substitution instances of '$\exists z(Rzj)$'; the instantiating constants are 'a,' 'd,' and 'j,' respectively.

2. Universal elimination

If you have '$\forall x(Ax)$,' it is legitimate to infer that anything is an A. You can infer 'Aa,' 'Ab,' 'Az,' 'Ad_3.' This is, you can infer any substitution instance—in short, you can infer '$A\mathcal{c}$' for any constant '$\mathcal{c}$.' This is the general form of the universal elimination rule ($\forall$E):

$$\begin{array}{l|ll} m & \forall\chi(\mathcal{A}) & \\ & \mathcal{A}\boxed{\chi\Rightarrow c} & \forall\text{E } m \end{array}$$

Remember that the box mark for a substitution instance is not a symbol of $\mathcal{L}_P$, so you cannot write it directly in a proof. Instead, you write the substituted sentence with the constant '$\mathcal{c}$' replacing all occurrences of the variable 'χ' in '$\mathcal{A}$.' For example:

1	$\forall x(Mx \to Rxd)$	
2	$Ma \to Rad$	$\forall$E 1
3	$Md \to Rdd$	$\forall$E 1

3. Existential introduction

When is it legitimate to infer '$\exists x(Ax)$'? If you know that something is an A—for instance, if you have 'Aa' available in the proof.

This is the existential introduction rule ($\exists$I):

m	$\mathcal{A}$	
	$\exists\chi(\mathcal{A}\boxed{\boxed{\chi\Rightarrow c}})$	$\exists$I m

It is important to notice that '$\mathcal{A}\boxed{\boxed{\chi\Rightarrow c}}$' is not necessarily a substitution instance. We write it with double boxes to show that the variable 'χ' does not need to replace all occurrences of the constant 'c.' You can decide which occurrences to replace and which to leave in place. For example:

1	$Ma \to Rad$	
2	$\exists x(Ma \to Rax)$	$\exists$I 1
3	$\exists x(Mx \to Rxd)$	$\exists$I 1
4	$\exists x(Mx \to Rad)$	$\exists$I 1
5	$\exists y\exists x(Mx \to Ryd)$	$\exists$I 4
6	$\exists z\exists y\exists x(Mx \to Ryz)$	$\exists$I 5

4. Universal introduction

A universal claim like '$\forall x(Px)$' would be proven if every substitution instance of it had been proven, if every line 'Pa,' 'Pb,' ... were available in a proof. Alas, there is no hope of proving *every* substitution instance. That would require proving 'Pa,' 'Pb,' ..., 'Pj_2,' ..., 'Ps_7,' ..., and so on to infinity. There are infinitely many constants in $\mathcal{L}_P$, and so this process would never come to an end.

Consider a simple argument: $\forall x(Mx)$, therefore $\forall y(My)$

It makes no difference to the meaning whether we use the variable 'x' or the variable 'y,' so this argument is obviously valid. Suppose we begin in this way:

1	$\forall x(Mx)$	want $\forall y(My)$
2	Ma	$\forall$E 1

We have derived 'Ma.' Nothing stops us from using the same justification to derive 'Mb,' ..., 'Mj_2,' ..., 'Ms_7,' ..., and so on until we run out of space or patience. We have effectively shown the way to prove 'Mc' for any constant c. From this, '$\forall y(My)$' follows.

1	$\forall x(Mx)$	
2	Ma	$\forall$E 1
3	$\forall y(My)$	$\forall$I 2

It is important here that 'a' was just some arbitrary constant. We had not made any special assumptions about it. If 'Ma' were a premise of the argument, then this would not show anything about *all* y. For example:

1	$\forall x(Rxa)$	
2	Raa	$\forall$E 1
3	$\forall y(Ryy)$	not allowed!

This is the schematic form of the universal introduction rule ($\forall$I):

m	$\mathcal{A}$	
	$\forall\chi(\mathcal{A}\boxed{c^* \Rightarrow \chi})$	$\forall$I m

* The constant c must not occur in any undischarged assumption.

Note that we can do this for any constant that does not occur in an undischarged assumption and for any variable.

Note also that the constant may not occur in any **undischarged** assumption, but it may occur as the assumption of a subproof that we have already closed. For example, we can prove '$\forall z(Dz \to Dz)$' without any premises.

1	Df	want Df
2	Df	R 1
3	$Df \to Df$	$\to$I 1–2
4	$\forall z(Dz \to Dz)$	$\forall$I 3

5. Existential elimination

A sentence with an existential quantifier tells us that there is *some* member of the UoD that satisfies a formula. For example, '$\exists x(Sx)$' tells us (roughly) that there is at least one S. It does not tell us *which* member of the UoD satisfies S, however. We cannot immediately conclude 'Sa,' 'Sf_{23},' or any other substitution instance. What can we do?

Suppose that we knew both '$\exists x(Sx)$' and '$\forall x(Sx \to Tx)$.' We could reason in this way:

> Since '$\exists x(Sx)$,' there is something that is an S. We do not know which constants refer to this thing, if any do, so call this thing 'Ishmael'. From '$\forall x(Sx \to Tx)$,' it follows that if Ishmael an S, then it is a T. Therefore, Ishmael is a T. Because Ishmael is a T, we know '$\exists x(Tx)$.'

In this paragraph, we introduced a name for the thing that is an S. We gave it an arbitrary name ('Ishmael') so that we could reason about it and derive some consequences from there being an S. Since 'Ishmael' is just a bogus name introduced for the purpose of the proof and not a genuine constant, we could not mention it in the conclusion. Yet we could derive something that does not mention Ishmael; namely, '$\exists x(Tx)$.' This does follow from the two premises.

We want the existential elimination rule to work in a similar way. Yet since English language worlds like 'Ishmael' are not symbols of $\mathcal{L}_P$, we cannot use them in formal proofs. Instead, we will use constants of $\mathcal{L}_P$ which do not otherwise appear in the proof.

A constant that is used to stand in for whatever it is that satisfies an existential claim is called a **proxy**. Reasoning with the proxy must all occur inside a subproof, and the proxy cannot be a constant that is doing work elsewhere in the proof.

This is the schematic form of the existential elimination rule (∃E):

$$
\begin{array}{l|ll}
m & \exists \mathcal{X}(\mathcal{A}) & \\
n & \quad \mid\ \mathcal{A}\ \boxed{c^* \Rightarrow \mathcal{X}} & \\
p & \quad \mid\ \mathcal{B} & \\
 & \mathcal{B} & \exists\text{E } m,\ n\text{–}p
\end{array}
$$

* The constant c must not appear outside the subproof.

Remember that the proxy constant cannot appear in $\mathcal{B}$, the sentence you prove using ∃E.

It would be enough to require that the proxy constant not appear in '$\exists \mathcal{X}(\mathcal{A})$,' in '$\mathcal{B}$,' or in any undischarged assumption. In recognition of the fact that it is just a place holder that we use inside the subproof, though, we require an entirely new constant which does not appear anywhere else in the proof.

With this rule, we can give a formal proof that '$\exists x(Sx)$' and '$\forall x(Sx \to Tx)$' together entail '$\exists x(Tx)$.'

$$
\begin{array}{l|ll}
1 & \exists x(Sx) & \\
2 & \forall x(Sx \to Tx) & \text{want } \exists x(Tx) \\
\hline
3 & \quad \mid\ Sa & \\
4 & \quad \mid\ Sa \to Ta & \forall\text{E } 2 \\
5 & \quad \mid\ Ta & \to\text{E } 3, 4 \\
6 & \quad \mid\ \exists x(Tx) & \exists\text{I } 5 \\
7 & \exists x(Tx) & \exists\text{E } 1, 3\text{–}6
\end{array}
$$

Notice that this has effectively the same structure as the English-language argument with which we began, except that the subproof uses the proxy constant 'a' rather than the bogus name 'Ishmael'.

6. Quantifier negation

When translating from English to $\mathcal{L}_P$, we noted that '$\neg\exists x(\neg\mathcal{A})$' is logically equivalent to '$\forall x(\mathcal{A})$.' In $\mathcal{L}_P$, they are provably equivalent. We can prove one half of the equivalence with a rather gruesome proof:

1	$\forall x(Ax)$	want $\neg\exists x(\neg Ax)$
2	$\quad \exists x(\neg Ax)$	for reductio
3	$\quad\quad \neg Ac$	for $\exists$E
4	$\quad\quad\quad \forall x(Ax)$	for reductio
5	$\quad\quad\quad Ac$	$\forall$E 1
6	$\quad\quad\quad \neg Ac$	R 3
7	$\quad\quad \neg\forall x(Ax)$	$\neg$I 4–6
8	$\quad \forall x(Ax)$	R 1
9	$\quad \neg\forall x(Ax)$	$\exists$E 3–7
10	$\neg\exists x(\neg Ax)$	$\neg$I 2–9

In order to show that the two sentences are genuinely equivalent, we need a second proof that assumes '$\neg\exists x(\neg\mathcal{A})$' and derives '$\forall x(\mathcal{A})$.' We leave that proof as an exercise for the reader.

It will often be useful to translate between quantifiers by adding or subtracting negations in this way, so we add two derived rules for this purpose. These rules are called *quantifier negation* (QN):

$$\begin{array}{l} \neg\forall\mathscr{x}(\mathcal{A}) \Longleftrightarrow \exists\mathscr{x}(\neg\mathcal{A}) \\ \neg\exists\mathscr{x}(\mathcal{A}) \Longleftrightarrow \forall\mathscr{x}(\neg\mathcal{A}) \quad \text{QN} \end{array}$$

Since QN is a replacement rule, it can be used on whole sentence or on subformulae.

7. Rules for identity

The identity predicate is not part of $\mathcal{L}_P$, but we add it when we need to symbolize certain sentences. For proofs involving identity, we add two rules of proof.

Suppose you know that many things that are true of a are also true of b. For example: '$(Aa \wedge Ab)$,' '$(Ba \wedge Bb)$,' '$(\neg Ca \wedge \neg Cb)$,' '$(Da \wedge Db)$,' '$(\neg Ea \wedge \neg Eb)$,' and so on. This would not be enough to justify the conclusion '$a = b$.' In general, there are no sentences in $\mathcal{L}_P$ that do not already contain the identity predicate that could justify the conclusion '$a = b$.' This means that the identity introduction rule will not justify '$a = b$' or any other identity claim containing two different constants.

However, it is always true that '$a = a$.' In general, no premises are required in order to conclude that something is identical to itself. So this will be the identity introduction rule, abbreviated =I:

$$\mathscr{c} = \mathscr{c} \qquad =\text{I}$$

Notice that the =I rule does not require referring to any prior lines of the proof. For any constant $\mathscr{c}$, you can write '$\mathscr{c} = \mathscr{c}$' on any point with only the =I rule as justification.

If you have shown that '$a = b$,' then anything that is true of a must also be true of b. For any symbolized sentence with 'a' in it, you can replace some or all of the occurrences of 'a' with 'b' and produce an equivalent sentence. For example, if you already know 'Raa,' then you are justified in concluding 'Rab,' 'Rba,' 'Rbb.' Recall that '$\mathcal{A}\boxed{\boxed{\mathscr{a} \Rightarrow \mathscr{b}}}$' is the sentence produced by replacing '$\mathscr{a}$' in '$\mathcal{A}$' with '$\mathscr{b}$.' This is not the same as a substitution instance, because '$\mathscr{b}$' may replace some or all occurrences of '$\mathscr{a}$'. The identity elimination rule (=E) justifies replacing terms with other terms that are identical to it:

m	$\mathscr{a} = \mathscr{b}$	
n	$\mathcal{A}$	
	$\mathcal{A}\boxed{\boxed{\mathscr{a} \Rightarrow \mathscr{b}}}$	=E m, n
	$\mathcal{A}\boxed{\boxed{\mathscr{b} \Rightarrow \mathscr{a}}}$	=E m, n

To see the rules in action, consider this proof:

1	$\forall x \forall y\ (x = y)$	
2	$\exists x (Bx)$	
3	$\forall x (Bx \to \neg Cx)$	want $\neg \exists x (Cx)$
4	$\quad$ | Be	
5	$\quad$ | $\forall y\ (e = y)$	$\forall$E 1
6	$\quad$ | $e = f$	$\forall$E 5
7	$\quad$ | Bf	=E 6, 4
8	$\quad$ | $Bf \to \neg Cf$	$\forall$E 3
9	$\quad$ | $\neg Cf$	$\to$E 8, 7
10	$\neg Cf$	$\exists$E 2, 4–9
11	$\forall x (\neg Cx)$	$\forall$I 10
12	$\neg \exists x (Cx)$	QN 11

8. Proof-theoretic concepts

We will use the symbol '$\vdash$' to indicate that a proof is possible. This symbol is called the **turnstile**. When we write '$\{\mathcal{A}_1, \mathcal{A}_2, \ldots\} \vdash \mathcal{B}$,' this means that it is possible to give a proof of $\mathcal{B}$ with '$\mathcal{A}_1$,' '$\mathcal{A}_2$,' ... as premises. With just one premise, we leave out the curly braces, so '$\mathcal{A} \vdash \mathcal{B}$' means that there is a proof of '$\mathcal{B}$' with '$\mathcal{A}$' as a premise. Naturally, '$\vdash \mathcal{B}$' means that there is a proof of '$\mathcal{B}$' that has no premises.

Often, logical proofs are called **derivations**. So '$\mathcal{A} \vdash \mathcal{B}$' can be read as '$\mathcal{B}$ is derivable from $\mathcal{A}$.' A **theorem** is a sentence that is derivable without any premises; i.e., $\mathcal{T}$ is a theorem if and only if '$\vdash \mathcal{T}$.'

It is not too hard to show that something is a theorem—you just have to give a proof of it. How could you show that something is *not* a theorem? If its negation is a theorem, then you could provide a proof. For example, it is easy to prove '$\neg(Pa \wedge \neg Pa)$,' which shows that '$Pa \wedge \neg Pa$' cannot be a theorem. For a sentence that is neither a theorem nor the negation of a theorem, however, there is no easy way to show this. You would have to demonstrate not just that certain proof strategies fail, but that no proof is possible.

Even if you fail in trying to prove a sentence in a thousand different ways, perhaps the proof is just too long and complex for you to make out.

Two sentences '$\mathcal{A}$' and '$\mathcal{B}$' are **provably equivalent** if and only if each can be derived from the other; i.e., '$\mathcal{A} \vdash \mathcal{B}$' and '$\mathcal{B} \vdash \mathcal{A}$.' It is relatively easy to show that two sentences are provably equivalent—it just requires a pair of proofs. Showing that sentences are *not* provably equivalent would be much harder. It would be just as hard as showing that a sentence is not a theorem. (In fact, these problems are interchangeable. Can you think of a sentence that would be a theorem if and only if '$\mathcal{A}$' and '$\mathcal{B}$' were provably equivalent?)

The set of sentences $\{\mathcal{A}_1, \mathcal{A}_2, \ldots\}$ is **provably inconsistent** if and only if a contradiction is derivable from it; i.e., for some sentence $\mathcal{B}$, $\{\mathcal{A}_1, \mathcal{A}_2, \ldots\} \vdash \mathcal{B}$ and $\{\mathcal{A}_1, \mathcal{A}_2, \ldots\} \vdash \neg\mathcal{B}$. It is easy to show that a set is provably inconsistent: you just need to assume the sentences in the set and prove a contradiction. Showing that a set is *not* provably inconsistent will be much harder. It would require more than just providing a proof or two; it would require showing that proofs of a certain kind are *impossible*.

Practice Exercises

⋆ **Part A**

Provide a justification (rule and line numbers) for each line of proof that requires one.

1	$\forall x \exists y(Rxy \lor Ryx)$	
2	$\forall x(\neg Rmx)$	
3	$\exists y(Rmy \lor Rym)$	
4		$(Rma \lor Ram)$
5		$\neg Rma$
6		Ram
7		$\exists x Rxm$
8	$\exists x(Rxm)$	

1	$\forall x(\exists y(Lxy) \to \forall z(Lzx))$
2	Lab
3	$\exists y(Lay) \to \forall z(Lza)$
4	$\exists y(Lay)$
5	$\forall z(Lza)$
6	Lca
7	$\exists y(Lcy) \to \forall z(Lzc)$
8	$\exists y(Lcy)$
9	$\forall z(Lzc)$
10	Lcc
11	$\forall x(Lxx)$

1	$\forall x(Jx \to Kx)$	
2	$\exists x \forall y(Lxy)$	
3	$\forall x(Jx)$	
4		$\forall y(Lay)$
5		Ja
6		$(Ja \to Ka)$
7		Ka
8		Laa
9		$(Ka \land Laa)$
10		$\exists x(Kx \land Lxx)$
11	$\exists x(Kx \land Lxx)$	

1		$\neg(\exists x(Mx) \lor \forall x(\neg Mx))$
2		$\neg\exists x(Mx) \land \neg\forall x(\neg Mx)$
3		$\neg\exists x(Mx)$
4		$\forall x(\neg Mx)$
5		$\neg\forall x(\neg Mx)$
6	$\exists x(Mx) \lor \forall x(\neg Mx)$	

⋆ **Part B**

Provide a proof of each claim.

1. $\vdash \forall x(Fx) \lor \neg\forall x(Fx)$
2. $\{\forall x(Mx \leftrightarrow Nx), Ma \land \exists x(Rxa)\} \vdash \exists x(Nx)$
3. $\{\forall x(\neg Mx \lor Ljx), \forall x(Bx \to Ljx), \forall x(Mx \lor Bx)\} \vdash \forall x(Ljx)$
4. $\forall x(Cx \land Dt) \vdash \forall x(Cx) \land Dt$
5. $\exists x(Cx \lor Dt) \vdash \exists x(Cx) \lor Dt$

Part C

Provide a proof of the argument about Billy on p. 70.

Part D

Look back at Part D on p. 66. Provide proofs to show that each of the argument forms is valid in $\mathcal{L}_P$.

Part E

Aristotle and his successors identified other syllogistic forms. Symbolize each of the following argument forms in $\mathcal{L}_P$ and add the additional assumptions 'There is an A' and 'There is a B.' Then prove that the supplemented arguments forms are valid in $\mathcal{L}_P$.

Darapti: All As are Bs. All As are Cs. Therefore, some B is C.
Felapton: No Bs are Cs. All As are Bs. Therefore, some A is not C.
Barbari: All Bs are Cs. All As are Bs. Therefore, some A is C.
Camestros: All Cs are Bs. No As are Bs. Therefore, some A is not C.
Celaront: No Bs are Cs. All As are Bs. Therefore, some A is not C.
Cesaro: No Cs are Bs. All As are Bs. Therefore, some A is not C.
Fapesmo: All Bs are Cs. No As are Bs. Therefore, some C is not A.

Part F

Provide a proof of each claim.

1. $\forall x \forall y(Gxy) \vdash \exists x(Gxx)$
2. $\forall x \forall y(Gxy \rightarrow Gyx) \vdash \forall x \forall y(Gxy \leftrightarrow Gyx)$
3. $\{\forall x(Ax \rightarrow Bx), \exists x(Ax)\} \vdash \exists x(Bx)$
4. $\{Na \rightarrow \forall x(Mx \leftrightarrow Ma), Ma, \neg Mb\} \vdash \neg Na$
5. $\vdash \forall z(Pz \vee \neg Pz)$
6. $\vdash \forall x(Rxx) \rightarrow \exists x \exists y(Rxy)$
7. $\vdash \forall y \exists x(Qy \rightarrow Qx)$

Part G

Show that each pair is provably equivalent.

1. $\forall x(Ax \rightarrow \neg Bx)$, $\neg \exists x(Ax \wedge Bx)$
2. $\forall x(\neg Ax \rightarrow Bd)$, $\forall x(Ax) \vee Bd$
3. $\exists x(Px) \rightarrow Qc$, $\forall x(Px \rightarrow Qc)$

Part H

Show that each of the following is provably inconsistent.

1. $\{(Sa \rightarrow Tm), (Tm \rightarrow Sa), (Tm \wedge \neg Sa)\}$
2. $\{\neg \exists x(Rxa), \forall x \forall y(Ryx)\}$
3. $\{\neg \exists x \exists y(Lxy), Laa\}$
4. $\{\forall x(Px \rightarrow Qx), \forall z(Pz \rightarrow Rz), \forall y(Py), \neg(Qa \wedge \neg Rb)\}$

⋆ Part I

Write a symbolization key for the following argument, translate it, and prove it:

> There is someone who likes everyone who likes everyone that he likes. Therefore, there is someone who likes himself.

Part J

Provide a proof of each claim.

1. $\{(Pa \vee Qb), (Qb \rightarrow b = c), \neg Pa\} \vdash Qc$

2. $\{(m = n \lor n = o), An\} \vdash (Am \lor Ao)$
3. $\{\forall x(x = m), Rma\} \vdash \exists x(Rxx)$
4. $\neg\exists x(x \neq m) \vdash \forall x\forall y(Px \to Py)$
5. $\forall x\forall y(Rxy \to x = y) \vdash (Rab \to Rba)$
6. $\{\exists x(Jx), \exists x(\neg Jx)\} \vdash \exists x\exists y(x \neq y)$
7. $\{\forall x(x = n \leftrightarrow Mx), \forall x(Ox \lor \neg Mx)\} \vdash On$
8. $\{\exists x(Dx), \forall x(x = p \leftrightarrow Dx)\} \vdash Dp$
9. $\{\exists x(Kx \land \forall y(Ky \to x = y) \land Bx), Kd\} \vdash Bd$
10. $\vdash Pa \to \forall x(Px \lor x \neq a)$

⋆ Part K

For each of the following pairs of sentences: if they are logically equivalent in $\mathcal{L}_P$, give proofs to show this.

1. $\forall x(Px) \to Qc$, $\forall x(Px \to Qc)$
2. $\forall x(Px) \land Qc$, $\forall x(Px \land Qc)$
3. $Qc \lor \exists x(Qx)$, $\exists x(Qc \lor Qx)$
4. $\forall x\forall y\forall z(Bxyz)$, $\forall x(Bxxx)$
5. $\forall x\forall y(Dxy)$, $\forall y\forall x(Dxy)$
6. $\exists x\forall y(Dxy)$, $\forall y\exists x(Dxy)$

⋆ Part L

For each of the following arguments: if it is valid in $\mathcal{L}_P$, give a proof.

1. $\forall x\exists y(Rxy)$, therefore $\exists y\forall x(Rxy)$
2. $\exists y\forall x(Rxy)$, therefore $\forall x\exists y(Rxy)$
3. $\exists x(Px \land \neg Qx)$, therefore $\forall x(Px \to \neg Qx)$
4. $\forall x(Sx \to Ta)$, Sd, therefore Ta
5. $\forall x(Ax \to Bx)$, $\forall x(Bx \to Cx)$, therefore $\forall x(Ax \to Cx)$
6. $\exists x(Dx \lor Ex)$, $\forall x(Dx \to Fx)$, therefore $\exists x(Dx \land Fx)$
7. $\forall x\forall y(Rxy \lor Ryx)$, therefore Rjj
8. $\exists x\exists y(Rxy \lor Ryx)$, therefore Rjj
9. $\forall x(Px) \to \forall x(Qx)$, $\exists x(\neg Px)$, therefore $\exists x(\neg Qx)$
10. $\exists x(Mx) \to \exists x(Nx)$, $\neg\exists x(Nx)$, therefore $\forall x(\neg Mx)$

Part M

1. If you know that $\mathcal{A} \vdash \mathcal{B}$, what can you say about $(\mathcal{A} \land \mathcal{C}) \vdash \mathcal{B}$? Explain your answer.
2. If you know that $\mathcal{A} \vdash \mathcal{B}$, what can you say about $(\mathcal{A} \lor \mathcal{C}) \vdash \mathcal{B}$? Explain your answer.

Appendix C

Solutions to Selected Exercises

Chapter 1 Part A

1. Declarative
2. Declarative
3. Not declarative
4. Declarative
5. Declarative
6. Declarative
7. Not declarative
8. Not declarative
9. Declarative
10. Not declarative

Chapter 1 Part B

1. True
2. True
3. True
4. False
5. True
6. False
7. True
8. False
9. False
10. True

Chapter 2 Part A

1. $\neg M$
2. $G \wedge \neg C$
3. $M \vee \neg M$
4. $G \vee C$
5. $\neg(G \vee C)$
6. $\neg M \wedge \neg G$
7. $(G \vee C) \wedge \neg(G \wedge C)$

Chapter 2 Part D
Symbolization key:

A : Alice lifted the couch.
B : Bob lifted the couch.
C : Alice and Bob together lifted the couch.

1. $A \wedge B$
2. $(A \vee B) \wedge \neg(A \wedge B)$
3. C
4. $\neg A \wedge B$
5. $\neg(A \vee B)$

Chapter 2 Part E
Possible symbolization key:

A : Alice is playing football.
C : Charlie is playing football.

Equivalent
1. Alice and Charlie are both playing football. $A \wedge C$
2. Both Alice and Charlie are playing football. $A \wedge C$

Non-equivalent
1. Alice and Charlie are not both playing football. $\neg(A \wedge C)$
2. Alice and Charlie are both not playing football. $\neg A \wedge \neg C$

Non-equivalent
1. Either Alice or Charlie is playing football. $A \vee C$
2. Neither Alice nor Charlie is playing football. $\neg(A \vee C)$, or $\neg A \wedge \neg C$

Equivalent
1. Neither Alice nor Charlie is playing football. $\neg(A \vee C)$, or $\neg A \wedge \neg C$
2. Alice is not playing football, and Charlie is not playing football either. $\neg A \wedge \neg C$

Non-equivalent
1. Either Alice is not playing football, or Charlie isn't. $\neg A \vee \neg C$
2. It's not the case that Alice or Charlie is playing football. $\neg(A \vee C)$

Chapter 3 Part B

1. $E_1 \wedge E_2$
2. $F_1 \rightarrow S_1$
3. $F_1 \vee E_1$
4. $E_2 \wedge \neg S_2$
5. $\neg E_1 \wedge \neg E_2$
6. $(E_1 \wedge E_2) \wedge \neg(S_1 \vee S_2)$
7. $S_2 \rightarrow F_2$

8. $(\neg E_1 \rightarrow \neg E_2) \wedge (E_1 \rightarrow E_2)$
9. $S_1 \leftrightarrow \neg S_2$
10. $(E_2 \wedge F_2) \rightarrow S_2$
11. $\neg(E_2 \wedge F_2)$
12. $(F_1 \wedge F_2) \leftrightarrow (\neg E_1 \wedge \neg E_2)$

Chapter 3 Part C

A : Alice is a spy.
B : Bob is a spy.
C : The code has been broken.
G : The German embassy will be in an uproar.

1. $A \wedge B$
2. $(A \vee B) \rightarrow C$
3. $\neg(A \vee B) \rightarrow \neg C$
4. $G \vee C$
5. $(C \vee \neg C) \wedge G$
6. $(A \vee B) \wedge \neg(A \wedge B)$

Chapter 4 Part B

1. tautology
2. contradiction
3. contingent
4. tautology
5. tautology
6. contingent
7. tautology
8. contradiction
9. tautology
10. contradiction
11. tautology
12. contingent
13. contradiction
14. contingent
15. tautology
16. tautology
17. contingent
18. contingent

Chapter 4 Part C 2, 3, 5, 6, 8, and 9 are logically equivalent.

Chapter 4 Part D 1, 3, 6, 7, and 8 are consistent.

Chapter 4 Part E 3, 5, 8, and 10 are valid.

Chapter 4 Part F

1. $\mathcal{X}$ and $\mathcal{Y}$ have the same truth value on every line of a complete truth table, so $\mathcal{X} \leftrightarrow \mathcal{Y}$ is true on every line. It is a tautology.

2. The proposition is false on some line of a complete truth table. On that line, $\mathcal{X}$ and $\mathcal{Y}$ are true and $\mathcal{C}$ is false. So the argument is invalid.
3. Since there is no line of a complete truth table on which all three propositions are true, the conjunction is false on every line. So it is a contradiction. The status of $\mathcal{Y} \wedge (\mathcal{X} \wedge \mathcal{C})$ is exactly the same: the order of the conjuncts makes no difference.
4. Since $\mathcal{X}$ is false on every line of a complete truth table, there is no line on which $\mathcal{X}$ and $\mathcal{Y}$ are true and $\mathcal{C}$ is false. So the argument is valid.
5. Since $\mathcal{C}$ is true on every line of a complete truth table, there is no line on which $\mathcal{X}$ and $\mathcal{Y}$ are true and $\mathcal{C}$ is false. So the argument is valid.
6. Not much. $\mathcal{X} \vee \mathcal{Y}$ is a tautology if $\mathcal{X}$ and $\mathcal{Y}$ are tautologies; it is a contradiction if they are contradictions; it is contingent if they are contingent.
7. $\mathcal{X}$ and $\mathcal{Y}$ have different truth values on at least one line of a complete truth table, and $\mathcal{X} \vee \mathcal{Y}$ will be true on that line. On other lines, it might be true or false. So $\mathcal{X} \vee \mathcal{Y}$ is either a tautology or it is contingent; it is *not* a contradiction.

Chapter 4 Part G

1. $\neg(\neg A \wedge \neg B)$
2. $\neg(A \wedge \neg B) \wedge \neg(\neg A \wedge B)$
3. $\neg A \vee B$
4. $\neg(\neg A \vee \neg B)$
5. $\neg(\neg A \vee \neg B) \vee \neg(A \vee B)$

Chapter 5 Part A

1. $Za \wedge (Zb \wedge Zc)$
2. $Rb \wedge \neg Ab$
3. $Lcb \rightarrow Mb$
4. $(Ab \wedge Ac) \rightarrow (Lab \wedge Lac)$
5. $\exists x(Rx \wedge Zx)$
6. $\forall x(Ax \rightarrow Rx)$
7. $\exists x(Rx \wedge \neg Ax)$
8. $\exists x(Rx \wedge Lcx)$
9. $\forall x((Mx \wedge Zx) \rightarrow Lbx))$

Chapter 5 Part B

1. $\neg\exists x(Tx)$
2. $\forall x(Mx \rightarrow Sx)$
3. $\exists x(\neg Sx)$
4. $\forall x(Sx \rightarrow Mx)$
5. $\neg\exists x(Bxx)$
6. $\neg\exists x(Cx \wedge (\neg Sx \wedge Tx))$

Chapter 5 Part B

1. $\forall x(Cxp \rightarrow Dx)$
2. $Cjp \wedge Fj$
3. $\exists x(Cxp \wedge Fx)$
4. $\neg\exists x(Sxj)$

5. $\forall x\big((Cxp \land Fx) \to Dx\big)$
6. $\neg\exists x(Cxp \land Mx)$
7. $\exists x(Cjx \land Sxe \land Fj)$
8. $Spe \land Mp$
9. $\forall x\big((Sxp \land Mx) \to \neg\exists y(Cyx)\big)$
10. $\exists x(Sxj \land \exists y(Cyx) \land Fj)$
11. $\forall x\big(Dx \to \exists y(Sxy \land Fy \land Dy)\big)$
12. $\forall x\big((Mx \land Dx) \to \exists y(Cxy \land Dy)\big)$

Chapter 5 Part D

1. $\forall x(Cx \to Bx)$
2. $\neg\exists x(Wx)$
3. $\exists x\exists y(Cx \land Cy \land x \neq y)$
4. $\exists x\exists y(Jx \land Ox \land Jy \land Oy \land x \neq y)$
5. $\forall x\forall y\forall z\big((Jx \land Ox \land Jy \land Oy \land Jz \land Oz) \to (x = y \lor x = z \lor y = z)\big)$
6. $\exists x\exists y\big(Jx \land Bx \land Jy \land By \land x \neq y \land \forall z((Jz \land Bz) \to (x = z \lor y = z))\big)$
7. $\exists x_1\exists x_2\exists x_3\exists x_4\big(Dx_1 \land Dx_2 \land Dx_3 \land Dx_4 \land x_1 \neq x_2 \land x_1 \neq x_3 \land x_1 \neq x_4 \land x_2 \neq x_3 \land x_2 \neq x_4 \land x_3 \neq x_4 \land \neg\exists y(Dy \land y \neq x_1 \land y \neq x_2 \land y \neq x_3 \land y \neq x_4)\big)$
8. $\exists x\big(Dx \land Cx \land \forall y((Dy \land Cy) \to x = y) \land Bx\big)$
9. $\forall x\big((Ox \land Jx) \to Wx\big) \land \exists x\big(Mx \land \forall y(My \to x = y) \land Wx\big)$
10. $\exists x\big(Dx \land Cx \land \forall y((Dy \land Cy) \to x = y) \land Wx\big) \to \exists x\forall y(Wx \leftrightarrow x = y)$
11. Wide scope: $\neg\exists x\big(Mx \land \forall y(My \to x = y) \land Jx\big)$
    Narrow scope: $\exists x\big(Mx \land \forall y(My \to x = y) \land \neg Jx\big)$
12. Wide scope: $\neg\exists x\exists z\big(Dx \land Cx \land Mz \land \forall y((Dy \land Cy) \to x = y) \land \forall y((My \to z = y) \land x = z)\big)$
    Narrow scope: $\exists x\exists z\big(Dx \land Cx \land Mz \land \forall y((Dy \land Cy) \to x = y) \land \forall y((My \to z = y) \land x \neq z)\big)$

Chapter 7 Part B

1. Yes: $pr(S \land V)$ is not 0, which it would be if S and V were inconsistent with one another.
2. No: $pr(S \land V)$ does not equal $pr(S)$ (or $pr(V)$), which it would if S and V were equivalent to one another.
3. $pr(S \lor V) = 0.375$.
4. $pr((S \lor V) \land S) = 0.25$.

Chapter 8 Part B

3, 5 and 6 are all cases of probabilistic independence.

Chapter 8 Part D

1. $pr(F) = 0.01$
2. $pr(\neg F) = 0.99$
3. $pr(T|F) = 0.99$
4. $pr(\neg T|\neg F) = 0.9$
5. $pr(T|\neg F) = 0.1$
6. $pr(T) = 0.1089$
7. $pr(\neg T) = 0.8911$
8. $pr(F|T) = 0.0909$
9. $pr(\neg F|\neg T) = 0.9999$

Chapter A Part A

1	$(W \to \neg B)$	
2	$(A \land W)$	
3	$(B \lor (J \land K))$	
4	W	$\land$E 2
5	$\neg B$	$\to$E 1, 4
6	$(J \land K)$	$\lor$E 3, 5
7	K	$\land$E 6

1	$(L \leftrightarrow \neg O)$	
2	$(L \lor \neg O)$	
3	$\quad \neg L$	
4	$\quad \neg O$	$\lor$E 2, 3
5	$\quad L$	$\leftrightarrow$E 1, 4
6	$\quad \neg L$	R 3
7	L	$\neg$E 3–6

1	$(Z \to (C \land \neg N))$	
2	$(\neg Z \to (N \land \neg C))$	
3	$\quad \neg(N \lor C)$	
4	$\quad (\neg N \land \neg C)$	DeM 3
5	$\quad\quad Z$	
6	$\quad\quad (C \land \neg N)$	$\to$E 1, 5
7	$\quad\quad C$	$\land$E 6
8	$\quad\quad \neg C$	$\land$E 4
9	$\quad \neg Z$	$\neg$I 5–8
10	$\quad (N \land \neg C)$	$\to$E 2, 9
11	$\quad N$	$\land$E 10
12	$\quad \neg N$	$\land$E 4
13	$(N \lor C)$	$\neg$E 3–12

Chapter A Part B

1.

1	$(K \land L)$	want $(K \leftrightarrow L)$
2	$\quad K$	want L
3	$\quad L$	$\land$E 1
4	$\quad L$	want K
5	$\quad K$	$\land$E 1
6	$(K \leftrightarrow L)$	$\leftrightarrow$I 2–3, 4–5

2.

1	$(A \to (B \to C))$	want $((A \land B) \to C)$
2	$\quad (A \land B)$	want C
3	$\quad A$	$\land$E 2
4	$\quad (B \to C)$	$\to$E 1, 3
5	$\quad B$	$\land$E 2
6	$\quad C$	$\to$E 4, 5
7	$((A \land B) \to C)$	$\to$I 2–6

3.

1	$(P \wedge (Q \vee R))$	
2	$(P \to \neg R)$	want $(Q \vee E)$
3	P	$\wedge$E 1
4	$\neg R$	$\to$E 2, 3
5	$(Q \vee R)$	$\wedge$E 1
6	Q	$\vee$E 5, 4
7	$(Q \vee E)$	$\vee$I 6

4.

1	$((C \wedge D) \vee E)$	want $E \vee D$
2	$\quad \neg E$	want D
3	$\quad (C \wedge D)$	$\vee$E 1, 2
4	$\quad D$	$\wedge$E 3
5	$(\neg E \to D)$	$\to$I 2–4
6	$(E \vee D)$	MC 5

5.

1	$(\neg F \to G)$	
2	$(F \to H)$	want $(G \vee H)$
3	$\quad \neg G$	want H
4	$\quad \neg\neg F$	MT 1, 3
5	$\quad F$	DN 4
6	$\quad H$	$\to$E 2, 5
7	$(\neg G \to H)$	$\to$I 3–6
8	$(G \vee H)$	MC 7

6.

1	$((X \wedge Y) \vee (X \wedge Z))$	
2	$\neg(X \wedge D)$	
3	$(D \vee M)$	want M
4	$\quad \neg X$	for reductio
5	$\quad (\neg X \vee \neg Y)$	$\vee$I 4
6	$\quad \neg(X \wedge Y)$	DeM 5
7	$\quad (X \wedge Z)$	$\vee$E 1, 6
8	$\quad X$	$\wedge$E 7
9	$\quad \neg X$	R 4
10	X	$\neg$E 4–9
11	$\quad \neg M$	for reductio
12	$\quad D$	$\vee$E 3, 11
13	$\quad (X \wedge D)$	$\wedge$I 10, 12
14	$\quad \neg(X \wedge D)$	R 2
15	M	$\neg$E 11–14

Chapter B Part A

1	$\forall x \exists y(Rxy \vee Ryx)$	
2	$\forall x(\neg Rmx)$	
3	$\exists y(Rmy \vee Rym)$	$\forall$E 1
4	$\quad (Rma \vee Ram)$	
5	$\quad \neg Rma$	$\forall$E 2
6	$\quad Ram$	$\vee$E 4, 5
7	$\quad \exists x Rxm$	$\exists$I 6
8	$\exists x(Rxm)$	$\exists$E 3, 4–7

1	$(\forall x(\exists y(Lxy) \to \forall z(Lzx)))$	
2	Lab	
3	$(\exists y(Lay) \to \forall z(Lza))$	$\forall$E 1
4	$\exists y(Lay)$	$\exists$I 2
5	$\forall z(Lza)$	$\to$E 3, 4
6	Lca	$\forall$E 5
7	$(\exists y(Lcy) \to \forall z(Lzc))$	$\forall$E 1
8	$\exists y(Lcy)$	$\exists$I 6
9	$\forall z(Lzc)$	$\to$E 7, 8
10	Lcc	$\forall$E 9
11	$\forall x(Lxx)$	$\forall$I 10

1	$\forall x(Jx \to Kx)$	
2	$\exists x \forall y(Lxy)$	
3	$\forall x(Jx)$	
4	$\quad \forall y(Lay)$	
5	$\quad Ja$	$\forall$E 3
6	$\quad (Ja \to Ka)$	$\forall$E 1
7	$\quad Ka$	$\to$E 6, 5
8	$\quad Laa$	$\forall$E 4
9	$\quad (Ka \wedge Laa)$	$\wedge$I 7, 8
10	$\quad \exists x(Kx \wedge Lxx)$	$\exists$I 9
11	$\exists x(Kx \wedge Lxx)$	$\exists$E 2, 4–10

1	$\quad (\neg(\exists x(Mx) \vee \forall x(\neg Mx)))$	
2	$\quad (\neg\exists x(Mx) \wedge \neg\forall x(\neg Mx))$	DeM 1
3	$\quad \neg\exists x(Mx)$	$\wedge$E 2
4	$\quad \forall x(\neg Mx)$	QN 3
5	$\quad \neg\forall x(\neg Mx)$	$\wedge$E 2
6	$(\exists x(Mx) \vee \forall x(\neg Mx))$	$\neg$E 1–5

Chapter B Part B

1.

1	$\quad\vert\; (\neg(\forall x(Fx) \vee \neg\forall x(Fx)))$	for reductio
2	$\quad\vert\; (\neg\forall x(Fx) \wedge \neg\neg\forall x(Fx))$	DeM 1
3	$\quad\vert\; \neg\forall x(Fx)$	$\wedge$E 2
4	$\quad\vert\; \neg\neg\forall x(Fx)$	$\wedge$E 2
5	$(\forall x(Fx) \vee \neg\forall x(Fx))$	$\neg$E 1–4

2.

1	$\forall x(Mx \leftrightarrow Nx)$	
2	$(Ma \wedge \exists x(Rxa))$	want $\exists x(Nx)$
3	$(Ma \leftrightarrow Na)$	$\forall$E 1
4	Ma	$\wedge$E 2
5	Na	$\leftrightarrow$E 3, 4
6	$\exists x(Nx)$	$\exists$I 5

3.

1	$\forall x(\neg Mx \vee Ljx)$	
2	$\forall x(Bx \rightarrow Ljx)$	
3	$\forall x(Mx \vee Bx)$	want $\forall x(Ljx)$
4	$(\neg Ma \vee Lja)$	$\forall$E 1
5	$(Ma \rightarrow Lja)$	MC 4
6	$(Ba \rightarrow Lja)$	$\forall$E 2
7	$(Ma \vee Ba)$	$\forall$E 3
8	Lja	DIL 7, 5, 6
9	$\forall x(Ljx)$	$\forall$I 8

4.

1	$\forall x(Cx \wedge Dt)$	want $(\forall x(Cx) \wedge Dt)$
2	$(Ca \wedge Dt)$	$\forall$E 1
3	Ca	$\wedge$E 2
4	$\forall x(Cx)$	$\forall$I 3
5	Dt	$\wedge$E 2
6	$(\forall x(Cx) \wedge Dt)$	$\wedge$I 4, 5

5.

1	$\exists x(Cx \vee Dt)$	want $(\exists x(Cx) \vee Dt)$
2	$\mid\ (Ca \vee Dt)$	for $\exists$E
3	$\mid\ \mid\ \neg(\exists x(Cx) \vee Dt)$	for reductio
4	$\mid\ \mid\ (\neg\exists x(Cx) \wedge \neg Dt)$	DeM 3
5	$\mid\ \mid\ \neg Dt$	$\wedge$E 4
6	$\mid\ \mid\ Ca$	$\vee$E 2, 5
7	$\mid\ \mid\ \exists x(Cx)$	$\exists$I 6
8	$\mid\ \mid\ \neg\exists x(Cx)$	$\wedge$E 4
9	$\mid\ (\exists x(Cx) \vee Dt)$	$\neg$E 3–8
10	$(\exists x(Cx) \vee Dt)$	$\exists$E 1, 2–9

Chapter B Part I Regarding the translation of this argument, see p. 72.

1	$\exists x\forall y(\forall z(Lxz \to Lyz) \to Lxy)$	
2	$\mid\ \forall y(\forall z(Laz \to Lyz) \to Lay)$	
3	$\mid\ (\forall z(Laz \to Laz) \to Laa)$	$\forall$E 2
4	$\mid\ \mid\ \neg\exists x(Lxx)$	for reductio
5	$\mid\ \mid\ \forall x(\neg Lxx)$	QN 4
6	$\mid\ \mid\ \neg Laa$	$\forall$E 5
7	$\mid\ \mid\ \neg\forall z(Laz \to Laz)$	MT 5, 6
8	$\mid\ \mid\ \mid\ Lab$	
9	$\mid\ \mid\ \mid\ Lab$	R 8
10	$\mid\ \mid\ (Lab \to Lab)$	$\to$I 8—9
11	$\mid\ \mid\ \forall z(Laz \to Laz)$	$\forall$I 10
12	$\mid\ \mid\ \neg\forall z(Laz \to Laz)$	R 7
13	$\mid\ \exists x(Lxx)$	$\neg$E 4—12
14	$\exists x(Lxx)$	$\exists$E 1, 2—13

Chapter B Part K 2, 3, and 5 are logically equivalent.

Chapter B Part L 2, 4, 5, 7, and 10 are valid. Here is a complete proof of 2:

1	$\exists y \forall x (Rxy)$	want $\forall x \exists y (Rxy)$
2	$\quad \forall x (Rxa)$	
3	$\quad Rba$	$\forall$E 2
4	$\quad \exists y (Rby)$	$\exists$I 3
5	$\quad \forall x \exists y (Rxy)$	$\forall$I 4
6	$\forall x \exists y (Rxy)$	$\exists$E 1, 2–5