

Logic 2: Modal Logic

Wolfgang Schwarz

January 27, 2024

© 2024 Wolfgang Schwarz

github.com/wo/logic2



This work is licensed under a [Creative Commons](https://creativecommons.org/licenses/by-nc-sa/4.0/) “Attribution-NonCommercial-ShareAlike 4.0 International” license.

Contents

Preface	3
1 Modal Operators	5
1.1 A new language	5
1.2 Flavours of modality	9
1.3 The turnstile	12
1.4 Duality	15
1.5 A system of modal logic	20

Preface

These notes are aimed at philosophy students who have taken an introductory course in formal logic. They provide an introduction to modal logic, with many philosophical applications. Along the way, they introduce general ideas that might be taught in an intermediate logic course: different methods of proof, the concept of a model, soundness and completeness, compactness, three-valued logics, free logics, supervaluation, properties of relations and orders, etc.

Chapters 1–3 introduce the standard toolkit of modal propositional logic: Kripke models, frame correspondence, some popular systems, the tableau method and axiomatic calculi. Chapter 4 goes through soundness and completeness. Chapters 5–8 turn to philosophical applications. Each of these chapters also extends the toolkit from chapter 3. Chapter 5 introduces multi-modal logics, chapter 6 ordering models and neighbourhood semantics, chapter 6 two-dimensional semantics and supervaluationism, chapter 7 conditional logics and Lewis-Stalnaker models. Chapters 9 and 10 look at some of the complexities that arise in first-order modal logic.

Apart from chapter 9, which sets the stage for chapter 10, every chapter after chapter 3 can be skipped or skimmed without affecting the accessibility of later chapters.

The best way to learn logic is by solving problems. That's why the text is frequently interrupted by exercises. As a student, you should try to do the exercises as soon as you reach them, before continuing with the text.

1 Modal Operators

1.1 A new language

Modal logic is an extension of propositional and predicate logic that is widely used to reason about possibility and necessity, obligation and permission, the flow of time, the processing of computer programs, and a range of other topics. Each of these applications begins by adding new symbols to the formal language of classical propositional or predicate logic. Before we explore such additions, let's briefly review why we use formal languages in the first place.

When reasoning about a given topic, we sometimes want to make sure that the stated conclusions really follow from the stated premises. If they do, we say that the reasoning is *valid*. By this we mean that there is no conceivable scenario in which the premises are true while the conclusions are false.

Here is an example of a valid argument.

All myriapods are oviparous.
Some arthropods are myriapods.
Therefore: Some arthropods are oviparous.

You can tell that this argument is valid even if you don't understand the zoological terms, because every argument of the same *logical form* is valid. The relevant logical form might be expressed as follows.

All F are G .
Some H are F .
Therefore: Some H are G .

No matter what descriptive terms you plug in for F , G , and H , you get a valid argument. The argument about myriapods is therefore not just valid, but *logically valid* – valid in virtue of its logical form.

In natural languages like English, the logical form of sentences is not always transparent. ‘Every dog barked at a tree’ can mean either that there is a single tree at which every dog barked, or that for each dog there is a tree at which it barked. The two readings have different logical consequences, so it would be good to keep them apart. Worse, the meaning of logical expressions (‘all’, ‘some’, ‘and’, etc.) in natural language is often unclear and complicated. ‘Paul and Paula got married and had children’ suggests that the marriage came before the children. In ‘Paul went to the zoo and Paula stayed at home’, the word ‘and’ does not seem to have this temporal meaning.

To get around these problems, we invent formal languages in which there are no ambiguities of logical form and in which all logical expressions have determinate, precise meanings. If we want to evaluate natural-language arguments for logical validity, we first have to translate them into the formal language. (Sometimes an argument will be valid on one translation and invalid on another.) With some practice, one can also reason directly in a formal language.

Now consider the following argument.

It might be raining.
It is certain that we will get wet if it is raining.
Therefore: We might get wet.

The argument looks valid. Indeed, any argument of this form is plausibly valid:

It might be that A .
It is certain that B if A .
Therefore: It might be that B .

But it’s hard to bring out the validity of these arguments in classical propositional or predicate logic. We need formal expressions corresponding to ‘it might be that’ and ‘it is certain that’. The languages of classical logic do not have such expressions.

So let’s add them. Let’s invent a new formal language with two new logical symbols. It doesn’t matter what these look like; a popular choice is a diamond \Diamond and a box \Box . We use the diamond to formalize ‘it might be that’, and the box for ‘it is certain that’.

If we add these symbols to the language of propositional logic, we get the standard language of modal propositional logic. If we add them to the language of predicate

logic, we get the standard language of modal predicate logic. We will stick with propositional logics until chapter ??.

Let's officially define the standard language of modal propositional logic.

Definition 1.1: The language \mathcal{L}_M

A *sentence letter* of \mathcal{L}_M is any lower-case letter of the Latin alphabet (a, b, c, \dots, z), possibly followed by numerical subscripts (a_1, p_{18}, \dots).
A *sentence* of \mathcal{L}_M is either a sentence letter of \mathcal{L}_M or an expression of the form $\neg A$, $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$, $(A \leftrightarrow B)$, $\Box A$, or $\Diamond A$, where A and B are \mathcal{L}_M -sentences.

I use lower-case letters a, b, c, \dots as atomic \mathcal{L}_M -sentences and upper-case letters A, B, C, \dots when I want to talk about arbitrary \mathcal{L}_M -sentences. To reduce clutter, I generally omit outermost parentheses and quotation marks when I mention \mathcal{L}_M -symbols or sentences: $p \wedge q$ is treated as an abbreviation of ' $(p \wedge q)$ '.

Exercise 1.1

Which of these are \mathcal{L}_M -sentences?

- (a) p
- (b) \Diamond
- (c) $\Diamond p \vee (\Box p \rightarrow p)$
- (d) $\Box \Box p$
- (e) $\Box A \rightarrow A$
- (f) $(\Diamond r \wedge \Diamond qr) \wedge \Diamond \Box \Diamond \Box p$

Having new symbols is only the beginning. We also need to lay down rules for reasoning with these symbols. The rules should be motivated by what the symbols are supposed to mean. So we shall also assign a more precise meaning to the diamond and the box – just as classical logic assigns a precise meaning to the symbol \wedge that may or may not exactly match the meaning of 'and' in English.

The meaning of \wedge can be given by a *truth table*:

A	B	$A \wedge B$
T	T	T
T	F	F
F	T	F
F	F	F

This tells us how the truth-value of $A \wedge B$ depends on the truth-value of A and B : the compound sentence is true iff (if and only if) both of its subsentences are true. If you know this, you know all there is to know about the meaning of \wedge . (You can see, for example, that $A \wedge B$ does not imply anything about the temporal order of A and B .)

Exercise 1.2

Draw the truth tables for \neg , \vee , \rightarrow , and \leftrightarrow .

The sentence operators (or connectives) of classical propositional logic (\neg , \wedge , \vee , \rightarrow , and \leftrightarrow) are all truth-functional. Recall that an operator is **truth-functional** if the truth-value of a compound sentence formed by applying the operator to other sentences is always determined by the truth-value of these other sentences. The truth tables for the classical operators spell out this dependence. They tell us how to compute the truth-value of a compound sentence from the truth-values of its constituents.

The diamond operator can't be truth-functional if it is supposed to mean anything like 'it might be that' in English. To see why, note first that 'it might be that P ' can be true if P is true, but also if P is false. 'It might be raining' doesn't entail that it is actually raining, nor that it isn't raining. It merely says that our evidence is compatible with rain. Now, if the diamond were truth-functional, then what would follow from the fact that $\Diamond p$ is *sometimes* true when p is true? It would follow that $\Diamond p$ is *always* true when p is true. (Make sure you understand why.) Likewise, from the fact that $\Diamond p$ is sometimes true when p is false, it would follow that $\Diamond p$ is true whenever p is false. $\Diamond p$ would be a logical truth. But 'it might be raining' is surely not a logical truth.

If an operator isn't truth-functional, its meaning can't be defined by a truth table. The standard approach to defining the meaning of modal operators instead involves the concept of possible worlds. Roughly, we'll interpret $\Diamond A$ as saying that A is true at some possible world, and $\Box A$ as saying that A is true at all possible worlds. Much

more on this later.

Exercise 1.3

Which of these English expressions are truth-functional?

- (a) It used to be the case that ...
- (b) It is widely known that ...
- (c) It is false that ...
- (d) It is necessary that ...
- (e) I can see that ...
- (f) God believes that ...
- (g) Either $2+2=4$ or it is practically feasible that ...

1.2 Flavours of modality

‘It might be that’ and ‘it is certain that’ express an *epistemic* kind of possibility and necessity, related to evidence and knowledge. There are other kinds – or *flavours* – of possibility and necessity.

Consider ‘John must leave’. This expresses a kind of necessity, but it would typically not be understood as a statement about the available evidence. On its most natural interpretation, it says that some relevant norms require John to leave. This flavour of necessity is called *deontic* (from Greek *deontos*: ‘of that which is binding’).

Other statements about possibility and necessity are neither deontic nor epistemic. If I say that you can’t travel from Auckland to Sydney by train, I don’t just mean that my information implies that you won’t make that journey; nor do I mean that you’re not permitted to make it. Rather, I mean that relevant circumstances in the world – such as the presence of an ocean between Auckland and Sydney – preclude the journey. This flavour of modality is sometimes called *circumstantial*. It comes in many sub-flavours, depending on what kinds of circumstances are in play.

Each of these flavours of modality corresponds to a branch of modal logic. *Epistemic logic* formalizes reasoning about knowledge and information. *Deontic logic* deals with norms, permissions, and obligations. A third branch of modal logic might be called *circumstantial logic*, but nobody uses that label. Some authors speak of

alethic modal logic (from *aletheia*: ‘truth’), but this label is also not used widely, and it is used for different things by different authors.

Confusingly, some philosophers use ‘modal logic’ for the logic of a certain sub-flavour of circumstantial modality, known as *metaphysical* modality. Metaphysical modality is concerned with what is or isn’t compatible with the nature of things. We will follow the more common practice of using ‘modal logic’ as an umbrella term that covers all the applications I have mentioned, as well as many others.

We will take a closer look at epistemic logic in chapter ?? and at deontic logic in chapter ?. In chapter ? we are going to study a branch of modal logic called *temporal logic* that is concerned with reasoning about time. Chapter ? is on *conditional logic*. Here we will introduce (non-truth-functional) two-place operators that are meant to formalise certain ‘if ...then ...’ constructions in English. In chapter ?, we will briefly look at *provability logic*, which investigates formal properties of mathematical provability. What unifies the different branches of modal logic is not a particular subject matter, but a loosely defined collection of abstract ideas and techniques that turn out to be useful in all these applications.

When we study some flavour of possibility or necessity, the diamond \Diamond is generally used for the relevant kind of possibility and the box \Box for the corresponding kind of necessity. In this context, you may pronounce the diamond ‘it is possible that’ and the box ‘it is necessary that’. In general, however, I would recommend pronouncing the diamond ‘diamond’ and the box ‘box’.

Different interpretations of the box and the diamond often motivate different rules for reasoning with these expressions. Consider, for example, the inference from $\Box p$ to p . If the box expresses a circumstantial kind of necessity, then this inference is plausibly valid: if the circumstances ensure that something is the case, then it really is the case. On a deontic reading of the box, by contrast, the inference is invalid. We can easily imagine scenarios in which, say, it is required that all library books are returned on time ($\Box p$) and yet it is not the case that all library books are returned on time ($\neg p$).

So we can’t say, once and for all, whether $\Box p$ entails p . We will develop different “logics” or “systems” of modal logic. In some systems, the inference is valid, in others it is invalid.

The diamond and the box are sentence operators. English expressions for necessity and possibility often don’t have this form. We can talk about what’s necessary or possible using ‘must’, ‘might’, or ‘can’, which are (auxiliary) verbs. We can also

use adjectives like ‘feasible’, ‘certain’, and ‘obligatory’, or adverbs like ‘possibly’, ‘certainly’, and ‘inevitably’.

When translating from English into \mathcal{L}_M , it is often helpful to first paraphrase the English sentence with ‘it is necessary that’ and ‘it is possible that’ (or other suitable sentence operators). For example,

You can’t go from Auckland to Sydney by train

might be paraphrased as

It is not possible [in light of relevant circumstances] that you go from
Auckland to Sydney by train

An adequate translation is $\neg\Diamond p$, where p represents ‘you go from Auckland to Sydney by train’ and the diamond represents the relevant kind of circumstantial possibility.

Exercise 1.4

Translate the following sentences, as well as possible, into \mathcal{L}_M , assuming that the diamond expresses epistemic possibility (‘it might be that’) and the box expresses epistemic necessity (‘it must be that’).

- (a) I may have offended the principal.
- (b) It can’t be raining.
- (c) Perhaps there is life on Mars.
- (d) If the murderer escaped through the window, there must be traces on the ground.
- (e) If the murderer escaped through the window, there might be traces on the ground.

Exercise 1.5

Translate the following sentences, as well as possible, into L_M , assuming that the diamond expresses deontic possibility (‘it is permitted that’) and the box expresses deontic necessity (‘it is obligatory that’).

- (a) I must go home.
- (b) You don’t have to come.

- (c) You can't have another beer.
- (d) If you don't have a ticket, you must pay a fine.

Exercise 1.6

Translate the following sentences, as well as possible, into L_M , assuming that the diamond expresses (some relevant sub-flavour of) circumstantial possibility and the box circumstantial necessity.

- (a) I could have studied architecture.
- (b) The bridge is fragile.
- (c) I can't hear you if you're talking to me from the kitchen.
- (d) If you have a smartphone, you can use an electronic ticket.

Special care is required when translating English sentences that contain both modal expressions and an 'if' clause. The surface form of English can be misleading. A good strategy is to first rephrase the English sentence so that it no longer contains any conditional expression, then translate that paraphrase. The paraphrase, and therefore the translation, will often sound rather unlike the original sentence, but that's OK. What's important is that it has the same truth-conditions. There should be no conceivable scenario in which the original sentence is true and the paraphrase (or translation) false, or the other way round.

1.3 The turnstile

In section 1.1, I said that an argument is valid if there is no conceivable scenario in which the premises are true and the conclusion is false. An argument is logically valid, I said, if it is valid "in virtue of its logical form". Can we make this more precise?

Consider this English argument.

Some cats are black.

Therefore: Some animals are black.

The argument is valid, but not logically valid. Its validity turns on the meaning of 'cat', which we don't consider a logical expression.

To bring out how the argument's validity depends on the meaning of 'cat', we can imagine a language that is much like English except that 'cat' means *chair*. In this language, the argument just displayed is invalid. It is invalid because there are conceivable scenarios in which there are black chairs but no black animals. In any such scenario, the argument's premise is true (in our imaginary language) while the conclusion is false.

When we say that an argument is valid "in virtue of its logical form", we mean that its validity does not depend on the meaning of the non-logical expressions. In other words, there is no conceivable scenario in which the premises are true and the conclusion is false, *no matter what meaning we assign to the non-logical expressions*.

The concept of validity for arguments is closely related to that of entailment. If an argument is valid, we say that the premises entail the conclusion. If an argument is logically valid, we say that the premises logically entail the conclusion. In logic, we're interested in logical entailment. We adopt the following definition.

Definition 1.2

Some sentences Γ ('gamma') (**logically**) **entail** a sentence A iff there is no conceivable scenario in which all sentences in Γ are true and A is false, under any interpretation of the non-logical expressions.

Instead of saying that the sentences Γ logically entail A , we also say that A is a *logical consequence* of Γ , or that A *logically follows from* Γ . Two sentences are (*logically*) *equivalent* if either logically follows from the other.

Logicians often use the symbol ' \models ' (the "double-barred turnstile") for entailment. The claim that $\Box(p \rightarrow q)$ and $\Box p$ together entail q , for example, could be expressed as

$$\Box(p \rightarrow q), \Box p \models q.$$

This is not a sentence of \mathcal{L}_M . The comma and the turnstile belong to the **meta-language** we use to talk about the **object language** \mathcal{L}_M . (The rest of our meta-language is mostly English.) We use the turnstile to express a certain relationship between \mathcal{L}_M -sentences, not to construct further \mathcal{L}_M -sentences.

Exercise 1.7

What do you think of this simpler alternative to definition 1.2? “Sentences Γ entail a sentence A iff there is no interpretation of non-logical expressions that renders all sentences in Γ true and A false.”

The following fact about logical consequence often proves useful.

Observation 1.1: If A and B are sentences and Γ is a (possibly empty) list of sentences, then

$$\Gamma, A \models B \text{ iff } \Gamma \models A \rightarrow B.$$

Proof. Look at the statement on the right-hand side of the ‘iff’. ‘ $\Gamma \models A \rightarrow B$ ’ says that there is no conceivable scenario in which all sentences in Γ are true while $A \rightarrow B$ is false, under any interpretation of the non-logical expressions. By the truth-table for ‘ \rightarrow ’, $A \rightarrow B$ is false iff A is true and B is false. So we can rephrase the statement on the right-hand side as saying that there is no conceivable scenario and interpretation that makes all sentences in Γ true and A true and B false. That’s just what the statement on the left-hand side asserts. \square

Observation 1.1 tells us that if we start with a claim of the form $A_1, A_2, A_3 \dots \models B$, we can always generate an equivalent claim by moving the turnstile to the left of the sentence that precedes it and putting an arrow in its original place. For example, instead of

$$\Box(p \rightarrow q), \Box p \models \Box q$$

we can equivalently say

$$\Box(p \rightarrow q) \models \Box p \rightarrow \Box q.$$

We can go further to

$$\models \Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q).$$

This says that $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$ logically follows from no premises at all. A sentence that follows from no premises is called *logically true* or (*logically*) *valid*.

(So an *argument* is called valid if the conclusion follows from the premises, while a *sentence* is called valid if it follows from no premises.)

Sentence validity is implicitly covered by definition 1.2, using an empty list of sentences for Γ . But it's worth making the definition more explicit.

Definition 1.3

A sentence A is **valid** (for short, $\models A$) iff there is no conceivable scenario in which A is false, under any interpretation of the non-logical expressions.

Make sure you don't confuse the arrow with the turnstile. It's not just that the two symbols belong to different languages – one to \mathcal{L}_M , the other to our meta-language. They also have very different meanings. $p \rightarrow q$ is true iff either p is false or q is true (or both). $p \models q$, on the other hand, is true iff there is no conceivable scenario in which p is true and q is false, under any interpretation of p and q . Nonetheless, there is an important connection between the arrow and the turnstile: $A \models B$ is *true* iff $A \rightarrow B$ is *valid*.

The definitions of this section are still somewhat imprecise. Eventually we will want to prove various claims about entailment and validity. To this end, we will need to give rigorous meanings to 'conceivable scenario' and 'interpretation of non-logical expressions'. Let's leave this task until the next chapter.

1.4 Duality

'Neville can't be the murderer', says Watson. His claim could be paraphrased as 'it is not possible that Neville is the murderer'. This suggests that $\neg\Diamond p$ is an adequate translation (where p expresses that Neville is the murderer). But Watson's claim might also be paraphrased as 'it is certain that Neville is not the murderer', which we might translate as $\Box\neg p$.

The two paraphrases are plausibly equivalent. In general, 'it is not (epistemically) possible that A ' seems to say the same as 'it is certain that not A '. Similarly, 'it is not certain that A ' arguably says the same as 'it is possible that not A '.

Whether or not the equivalence holds in English, we stipulate that it holds in \mathcal{E}_M : for any \mathcal{E}_M -sentence A ,

(Dual1) $\neg\Diamond A$ is equivalent to $\Box\neg A$;

(Dual2) $\neg\Box A$ is equivalent to $\Diamond\neg A$.

Operators that stand in the relationship expressed by (Dual1) and (Dual2) are called **duals** of each other. There is a convention in modal logic to use the symbols \Box and \Diamond only for concepts that are duals of each other.

Exercise 1.8

Find all pairs of duals among the following English expressions.

- (a) It is necessary that ...
- (b) It is impossible that ...
- (c) It is possible that ...
- (d) It is possibly not the case that ...
- (e) It was at some point the case that ...
- (f) It will at some point be the case that ...
- (g) It has always been the case that ...
- (h) It will always be the case that ...
- (i) The law requires that ...
- (j) The law does not require that ...
- (k) The law allows that ...
- (l) It is true that ...
- (m) It is false that ...

(Dual1) implies that $\neg\Diamond\neg p$ is equivalent to $\Box\neg\neg p$, choosing $\neg p$ as the sentence A . In standard modal logic, logically equivalent expressions are interchangeable. So we can simplify $\Box\neg\neg p$ to $\Box p$, drawing on the equivalence between $\neg\neg p$ and p . So $\neg\Diamond\neg p$ is equivalent to $\Box p$.

The same reasoning could be applied to any other sentence A in place of p . (Dual1) therefore implies that for any sentence A ,

$$\Box A \text{ is equivalent to } \neg\Diamond\neg A.$$

In the same way, (Dual2) implies that (for any sentence A)

$$\Diamond A \text{ is equivalent to } \neg \Box \neg A.$$

This shows that the box and the diamond can be defined in terms of one another. We could have used a language whose only primitive modal operator is the box, and read $\Diamond A$ as an abbreviation of $\neg \Box \neg A$. Alternatively, we could have used the diamond as the only primitive modal operator and read $\Box A$ as an abbreviation of $\neg \Diamond \neg A$.

Exercise 1.9

Which of these sentences are equivalent to $\Diamond \Diamond \neg p$? (a) $\Diamond \neg \Diamond p$, (b) $\Diamond \neg \Box p$, (c) $\neg \Box \Diamond p$, (d) $\neg \Diamond \Box p$, (e) $\neg \Box \Box p$

A digression: you might think that there is another connection between ‘possible’ and ‘necessary’. When we say that something is possible (or that it might be the case), we often convey that it is not necessary (or not certain). This suggests that $\Diamond p$ entails $\neg \Box p$. We’ve just assumed, however, that $\Diamond p$ is equivalent to $\neg \Box \neg p$. If $\Diamond p$ entails $\neg \Box p$, we would have to conclude that $\neg \Box \neg p$ entails $\neg \Box p$. By contraposition, we could infer that $\Box p$ entails $\Box \neg p$. But ‘it is necessary that P ’ surely doesn’t entail ‘it is necessary that not- P ’!

We have to reject either the duality of ‘possible’ and ‘necessary’ or the apparent entailment from ‘possible’ to ‘not necessary’. On reflection, the case for duality is stronger. There is a good explanation of why ‘possible’ often *appears* to entail ‘not necessary’ even if it actually doesn’t.

Take an example. Suppose Watson says ‘Neville might be the murderer’. Let’s assume that ‘might’ is the dual of ‘certain’, so that ‘it might be that P ’ is equivalent to ‘it is not certain that not P ’. On this interpretation, what Watson said – that Neville might be the murderer – is merely that it isn’t certain that Neville is *not* the murderer. It may well be certain that Neville *is* the murderer. Why, then, does his statement convey that Neville’s guilt is an open question?

Well, suppose Watson had known that Neville is the murderer. In that case, he shouldn’t have said ‘Neville might be the murderer’. These words would still have been true – or so we assume – but they would not have been helpful. Watson would have been in a position to say something more informative: that Neville is the murderer, or that he is known to be the murderer. We generally assume that speakers

are trying to be helpful, that they are not hiding relevant information. Assuming that Watson is trying to be helpful, his *statement* that Neville might be the murderer implies that he considers Neville's guilt an open question. This follows not from *what he said*, but from the fact *that he said it*, together with the assumption that he is trying to be helpful.

This kind of effect is studied in the field of pragmatics, where it is known as a *scalar implicature*. Scalar implicatures arise when an utterance of a logically weaker sentence conveys that a certain stronger sentence is false. 'Some students passed the test', for example, conveys that not all students passed the test, although the statement would be true even if all students had passed. In that case, however, it would not have been helpful: the speaker should have used 'all students passed'. End of digression.

I want to say a little more about duality. To do so, I need to introduce the concept of a schema.

Formally, a **schema** (for \mathcal{L}_M -sentences) is simply an \mathcal{L}_M -sentence with upper-case schematic variables in place of sentence letters. Every \mathcal{L}_M -sentence that results from a schema by (uniformly) replacing the schematic variables with object-language sentences is called an **instance** of the schema.

$\Box A \rightarrow A$, for example, is a schema. Three of its instances are $\Box p \rightarrow p$ and $\Box(p \vee q) \rightarrow (p \vee q)$ and $\Box\Box p \rightarrow \Box p$. The sentence $\Box p \rightarrow q$ is not an instance: the same schematic variable must always be replaced by the same object-language sentence. (That's what I meant by "uniformly".)

Exercise 1.10

Which of the following expressions are instances of $\Box(A \rightarrow \Diamond(A \wedge B))$?

- (a) $\Box(p \rightarrow \Diamond(q \wedge p))$
- (b) $\Box(\Diamond p \rightarrow \Diamond(\Diamond p \wedge p))$
- (c) $\Box\Box(p \rightarrow \Diamond(p \wedge q))$
- (d) $\Box((p \rightarrow \Diamond(p \wedge q)) \rightarrow \Diamond((p \rightarrow \Diamond(p \wedge q)) \wedge \Diamond p))$
- (e) $\Box((A \wedge C) \rightarrow \Diamond((A \wedge C) \wedge (B \wedge C)))$

Schemas are useful when we want to talk about all \mathcal{L}_M -sentences of a certain form. In the next section, for example, we are going to define a system of modal logic by giving a list of schemas all instances of which are considered valid.

Now compare the schemas $\Box A \rightarrow A$ and $A \rightarrow \Diamond A$. Given the duality of the box

and the diamond, and the fact that logically equivalent expressions can be freely exchanged for one another, we can show that *every instance of one of them is equivalent to an instance of the other*. In this sense, the two schemas are equivalent. And because their equivalence relies on the duality of the box and the diamond, the two schemas are called duals of one another.

To see why every instance of $\Box A \rightarrow A$ is equivalent to an instance of $A \rightarrow \Diamond A$, take a simple instance: $\Box p \rightarrow p$. By the truth-table for the arrow, this is equivalent to $\neg p \rightarrow \neg \Box p$. By (Dual2), $\neg \Box p$ is equivalent to $\Diamond \neg p$. So $\neg p \rightarrow \neg \Box p$ is equivalent to $\neg p \rightarrow \Diamond \neg p$. And this is an instance of $A \rightarrow \Diamond A$. The same line of reasoning obviously works for any other sentence in place of p , and a similar line of reasoning shows the converse, that every instance of $A \rightarrow \Diamond A$ is equivalent to an instance of $\Box A \rightarrow A$.

It's crucial that we're talking about schemas here. We have not shown that the *sentence* $\Box p \rightarrow p$ is equivalent to $p \rightarrow \Diamond p$. In fact, the duality principles and the replacement of equivalents don't suffice to show that these sentences are equivalent.

The equivalence of the *schemas*, however, is enough to show that it doesn't matter which of them we use when we list schemas to define a logic. We can say that all instances of $\Box A \rightarrow A$ are valid in a certain logic, or we can say that all instances of $A \rightarrow \Diamond A$ are valid – it amounts to the same thing, because every instance of either schema is equivalent to an instance of the other.

The equivalence between $\Box A \rightarrow A$ and $A \rightarrow \Diamond A$ is an example of a more general pattern. Any schema with an arrow (\rightarrow or \leftrightarrow) as the only truth-functional operator can be converted into an equivalent schema – its **dual** – by swapping antecedent and consequent and replacing every box with a diamond and every diamond with a box.

Exercise 1.11

Find the duals of (a) $\Box A \rightarrow \Box \Box A$, (b) $\Diamond A \rightarrow \Box \Diamond A$, (c) $\Box A \rightarrow \Diamond A$.

Exercise 1.12

A proposition is *contingent* if it is neither necessary nor impossible. Let ∇ be a sentence operator for 'it is contingent that'. Reading the box as 'it is necessary that' and the diamond as 'it is possible that', try to find

- (a) a sentence whose only modal operator is \Box that is equivalent to ∇p ;
- (b) a sentence whose only modal operator is \Diamond that is equivalent to ∇p ;

(c) a sentence whose only modal operator is ∇ that is equivalent to $\Box p$.

1.5 A system of modal logic

Whether a sentence is logically valid, or logically entailed by other sentences, never depends on the meaning of the non-logical expressions. But it may well depend on the meaning of the logical expressions. In modal logic, the box and the diamond are treated as logical expressions, but their interpretation varies from application to application. Sometimes the box means epistemic necessity, sometimes it means deontic necessity, sometimes it means something else. As I mentioned in section 1.2, this has the consequence that we need to distinguish different “systems of modal logic”. In some applications, we want $\Box p$ to entail p , in others we don’t.

Suppose, now, that we want to fully spell out one of these “systems”. We want to completely specify which \mathcal{L}_M -sentences are valid, and which are entailed by which others, on a particular understanding of the modal operators.

There are many ways of approaching this task. We could, for example, define precise notions of conceivable scenarios and interpretations and apply the definitions of the previous section. But let’s choose a more direct route. When we think about circumstantial necessity, we can intuitively see that $\Box p$ entails p , without going through sophisticated considerations about scenarios and interpretations. Assume, then, that we simply start with direct judgements about entailment and validity.

We still face a problem. There are infinitely many \mathcal{L}_M -sentences. We can’t look at every sentence and argument one by one. We need to find some shortcuts.

We can begin by drawing on a consequence of observation 1.1. Above I said that in order to spell out a system of modal logic, we need to specify (i) which \mathcal{L}_M -sentences are valid and (ii) which \mathcal{L}_M -sentences are entailed by which others. Observation 1.1 tells us that we can ignore part (ii) of the task. Once we have fixed which sentences are valid, we have implicitly also fixed which sentences entail which others. If, for example, we decide that $\Box p \rightarrow p$ is valid, we have also decided that $\Box p$ entails p .

Our task of spelling out a system of modal logic therefore reduces to the task of specifying which \mathcal{L}_M -sentences are valid. That’s why a **system of modal logic** is usually defined simply as a set of \mathcal{L}_M -sentences.

To make this more concrete, let’s look at a particular sub-flavour of circumstantial necessity, sometimes called *historical necessity*. Something is historically necessary

if it is “settled”: it is true and there is nothing anyone can do about it. Facts about the past are plausibly settled. Nothing we can do is going to make a difference to what happened yesterday. By contrast, some facts about the future are intuitively “open”.

Let’s use the box to formalise this (admittedly vague) concept of historical necessity. So $\Box p$ says that p is settled. Since the diamond is the dual of the box, $\Diamond p$ expresses that it not settled that p is false. In other words, p is either open or settled as true.

Our task is to specify all \mathcal{L}_M -sentences that are valid on this understanding of the box and the diamond. This will give us a system of modal logic, a set of \mathcal{L}_M -sentences that are valid on a certain interpretation of the box and the diamond. We want to know which sentences are in the system – for short, which sentences are “in” – and which are not.

If the box expresses historical necessity then $\Box p$ clearly entails p . So $\Box p \rightarrow p$ is in. There is nothing special here about the sentence p . Whatever is settled is true. Every instance of the schema $\Box A \rightarrow A$ is in. (As mentioned in section 1.4, it follows that every instance of $A \rightarrow \Diamond A$ is in as well.)

In the same vein, we may now look at other schemas. Arguably, all instances of the following schemas – listed here with their conventional names – are valid, and therefore in our target system:

- (Dual) $\neg \Diamond A \leftrightarrow \Box \neg A$
- (T) $\Box A \rightarrow A$
- (K) $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$
- (4) $\Box A \rightarrow \Box \Box A$
- (5) $\Diamond A \rightarrow \Box \Diamond A$

(Dual) corresponds to the duality principle (Dual1) from section 1.4. Its instances are guaranteed to be valid by the fact that we have introduced the diamond as the dual of the box.

We’ve already talked about (T).

(K) is a little easier to understand as a claim about entailment:

$$\Box(A \rightarrow B), \Box A \models \Box B.$$

On our present interpretation, this says that if a material conditional $A \rightarrow B$ is settled,

and its antecedent A is settled, then its consequent B is guaranteed to be settled as well. Why should we accept this? Let A and B be arbitrary propositions, and assume that $A \rightarrow B$ and A are both settled. It follows that they are both true. Since $A \rightarrow B$ and A entail B , it follows that B is true as well. Could it be that B is true but open? Arguably not: If we could bring about a situation in which B is false then we could also bring about a situation in which either $A \rightarrow B$ or A is false, since one of these is guaranteed to be false in any situation in which B is false. The assumption that $A \rightarrow B$ and A are settled therefore implies that B is settled. So all instances of (K) are in.

(4) and (5) assert that facts about what is settled are themselves settled. (4) says that if something is settled then it is settled that it is settled. (5) says that if something is not settled then it is settled that it is not settled. Here it is important that we adopt a consistent point of view. It is easy to think of situations in which something is open to us (say, we could read a certain letter) and we can do something (say, burn the letter) that would make it no longer open. This doesn't contradict (5), since (5) concerns what is open and settled *now*. If something is now open, then arguably there is nothing we can do that would change the fact that it is now open. Likewise, if something is now settled, then arguably there is nothing we can do that would change the fact that it is now settled.

I could have listed further schemas. For example, whenever a conjunction is settled, then both its conjuncts are plausibly settled as well. So every instance of $\Box(A \wedge B) \rightarrow (\Box A \wedge \Box B)$ should be in. There are, in fact, infinitely many further schemas, not covered by the five above, whose instances belong to our target system.

That's the bad news. The good news is that we don't need to list any of them. We can replace the whole lot by specifying two rules for generating new sentences from sentences we have already classified as "in".

The first of these rules captures the plausible thought that anything that follows from a valid sentence by classical (non-modal) propositional logic is itself valid. Since we've decided that $\Box p \rightarrow p$ is valid (in the logic of historical necessity), we can, for example, infer that $(\Box p \rightarrow p) \vee q$ is also valid, because $A \vee B$ follows from A in classical propositional logic. Our system of modal logic thereby becomes an **extension** of classical propositional logic.

To state the rule concisely, let $\Gamma \models_0 A$ mean that A follows from Γ in classical propositional logic – as can be determined, for example, by the truth table method.

Then our rule says that for any list of sentences Γ and any sentence A ,

(CPL) If $\Gamma \models_0 A$ and all members of Γ are in, then A is in.

As a special case, (CPL) implies that every propositional tautology is “in”, since tautologies follow in classical propositional logic from any premises whatsoever (and even from no premises).

Our second rule reflects the idea that all logical truths are settled: For any sentence A ,

(Nec) If A is in, then $\Box A$ is in.

And now we’re done. I claim – and this may seem rather mysterious at the moment – that there is a natural understanding of historical necessity (of ‘settled’) on which the sentences that are valid in the logic of historical necessity are precisely the sentences that can be generated from instances of (T), (K), (4), (5) and (Dual) by (CPL) and (Nec). (In fact, (4) is redundant: any instance of (4) can be derived from the remaining axioms and rules.)

The system of modal logic defined by these schemas and rules is perhaps the best known of all systems of modal logic. Its conventional name is ‘S5’ because it was introduced as the fifth system in an influential list of systems published by C.I. Lewis and C.H. Langford in 1932.

Other systems of modal logic can be defined by different schemas or rules. Lewis and Langford’s system S4, for example, is defined by (T), (K), (4), (Dual), (CPL) and (Nec), without (5). This system is adequate for other interpretations of the box and the diamond, where we don’t want to treat all instances of (5) as valid.

Exercise 1.13

Instead of reading the box as ‘it is settled that’, we might give it one of these interpretations (with the diamond defined as the box’s dual):

- (a) it is true that
- (b) it is false that
- (c) it is either true or false that
- (d) it is logically true that

For each of these interpretations, evaluate whether the schemas (T), (K), (4), (5), and the rules (CPL) and (Nec) are plausible.

Remember that a system of modal logic is just a set of \mathcal{L}_M -sentences. I have defined the system S5 in terms of (T), (K), (4), (5), or (Dual), (CPL) and (Nec), but the same system can be defined by many other combinations of schemas and rules. (Lewis and Langford used a very different definition.)

The schemas and rules that I have chosen are called an **axiomatisation** of S5. The schemas – or more precisely, their instances – are called **axioms** because they are the starting points if we want to show that a sentence is in the system.

To illustrate this point, think of how we could show that $\Box(p \wedge q) \rightarrow \Box p$ is in S5 (that it is “S5-valid”). The sentence is not an instance of any of the schemas I have listed. Instead, we may start with the non-modal sentence $(p \wedge q) \rightarrow p$. This is a propositional tautology, so (CPL) tells us that it is in S5. By (Nec), it follows that $\Box((p \wedge q) \rightarrow p)$ is in S5 as well. Since all instance of (K) are in S5, the system contains

$$\Box((p \wedge q) \rightarrow p) \rightarrow (\Box(p \wedge q) \rightarrow \Box p).$$

By Modus Ponens, $\Box((p \wedge q) \rightarrow p)$ and $\Box((p \wedge q) \rightarrow p) \rightarrow (\Box(p \wedge q) \rightarrow \Box p)$ entail our target sentence $\Box(p \wedge q) \rightarrow \Box p$. By (CPL), this means the target sentence is also in S5.

Here is a more streamlined presentation of this line of reasoning.

- | | |
|---|-------------|
| 1. $(p \wedge q) \rightarrow p$ | (CPL) |
| 2. $\Box((p \wedge q) \rightarrow p)$ | (1, Nec) |
| 3. $\Box((p \wedge q) \rightarrow p) \rightarrow (\Box(p \wedge q) \rightarrow \Box p)$ | (K) |
| 4. $\Box(p \wedge q) \rightarrow \Box p$ | (2, 3, CPL) |

We can use the same streamlined format to show that, say, $\Box p \rightarrow \Diamond p$ is S5-valid.

1. $\Box \neg p \rightarrow \neg p$ (T)
2. $\neg \Diamond p \leftrightarrow \Box \neg p$ (Dual)
3. $\neg \Diamond p \rightarrow \neg p$ (1, 2, CPL)
4. $p \rightarrow \Diamond p$ (3, CPL)
5. $\Box p \rightarrow p$ (T)
6. $\Box p \rightarrow \Diamond p$ (4, 5, CPL)

These annotated lists look a lot like proofs. They *are* proofs. Every axiomatisation of a logical system defines a corresponding **axiomatic calculus**. A proof in an axiomatic calculus is simply a list of sentences each of which is either an axiom or follows from earlier sentences in the list by one of the rules. (The annotations on the right are not officially part of the proof. They are added to help understand where the lines come from.)

Exercise 1.14

Try to find axiomatic proofs showing that the following sentences are in S5.

- (a) $\Box(\Box p \rightarrow p)$
- (b) $(\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)$
- (c) $\Diamond \neg p \leftrightarrow \neg \Box p$

Exercise 1.15

In the axiomatic calculus for S5, (Nec) allows us to derive $\Box A$ from A . Someone might object that this inference is obviously invalid, since a sentence might be true without being necessarily true. Can you explain why (Nec) is an acceptable rule in the axiomatic calculus for S5?

The axiomatic method is the oldest formal method of proof. It has many virtues, but user-friendliness is not among them. Even simple facts are often hard to prove in an axiomatic calculus. In the next chapter, we will meet a different method that is much easier to use.