1. Boxes and diamonds

1.1. Modal Operators

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*. This is usually cashed out as meaning 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 whose logical expressions have precise meanings and in which there are no ambiguities of logical form. 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.

Returning to modal logic, 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 to use a diamond $\Diamond$ and a box $\Box$. The diamond is meant to formalize ‘it might be that’, and the box ‘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 9.

We might translate the above argument into the language of modal propositional logic as follows, using r to mean that it is raining and w that we will get wet:

$$\begin{array}{l} \Diamond r \\ \Box(r \to w) \\ \hline \Diamond w \end{array}$$

Let's define exactly what counts as a sentence in the standard language of modal propositional logic, or $\mathfrak{L}_M$ for short.

Definition 1.1: The language $\mathfrak{L}_M$

1. Every sentence letter $p, q, r, \ldots$ is an $\mathfrak{L}_M$-sentence.
2. If A is an $\mathfrak{L}_M$-sentence, then so are $\neg A$, $\Diamond A$, and $\Box A$.
3. If A and B are $\mathfrak{L}_M$-sentences, then so are $(A \land B)$, $(A \lor B)$, $(A \to B)$ and $(A \leftrightarrow B)$.
4. Nothing else is an $\mathfrak{L}_M$-sentence.

Note that I use lower-case letters $p, q, r, \ldots$ for atomic sentences of $\mathfrak{L}_M$ and upper-case letters $A, B, C, \ldots$ when I want to talk about arbitrary $\mathfrak{L}_M$-sentences. To reduce clutter, I generally omit outermost parentheses and quotation marks when I mention $\mathfrak{L}_M$-sentences: $p \land q$ is treated as an abbreviation of '$(p \land q)$'.

Exercise 1.1

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

(a) p
(b) $\Diamond$
(c) $\Diamond p \lor (\Box p \to p)$
(d) $\Box\Box p$
(e) $\Box A \to A$
(f) $(\Diamond r \land \Diamond qr) \land \Diamond\Box\Diamond\Box p$

Of course, having new symbols is only the beginning. We will also need to lay down some new rules for how to reason 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 $\land$ that may or may not exactly match the meaning of 'and' in English.

We have entered the realm of modal logic.

Historically, modal logic grew out of the study of necessity and contingency, which medieval logicians regarded as "modes of truth". Hence the name 'modal logic'. Today, the study of necessity and contingency is but one of many subfields within modal logic. To a first (and rough) approximation, any part of logic that involves *non-truth-functional sentence operators* is part of modal logic.

By a **sentence operator** I mean an expression that combines with one or more sentences to create a new sentence. Classical propositional logic has the sentence operators $\neg$ (not), $\land$ (and), $\lor$ (or), $\to$ (if-then), and $\leftrightarrow$ (if-and-only-if). In modal logic, we have further operators such as $\Diamond$ and $\Box$. (Sentence operators are also called 'connectives'.)

The sentence operators of classical propositional logic are all truth-functional. Remember that an operator is **truth-functional** if the truth-value of a complex sentence built with the use the operator is determined by the truth-value of its parts. For example, a conjunction $A \land B$ is true whenever A and B are both true, and false otherwise. If you know the truth-value of the conjuncts, you can figure out the truth-value of the conjunction.

The meaning of truth-functional operators can be specified by a truth-table. Not so for our newly introduced operators $\Diamond$ and $\Box$. The standard approach to define 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 1.2

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 . . .

Exercise 1.3

Explain why the $\Diamond$ operator that formalises 'it might be that' is not truth-functional.

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. When we say that it might be raining, we convey that our evidence leaves open the possibility that it is raining. There are other kinds – often called *flavours* – of possibility and necessity.

Consider 'John must leave'. This would typically not be understood as a statement about the available evidence, but about contextually salient norms. The sentence says that *by these norms* it is necessary that John leaves. Here we are dealing with a *deontic* flavour of necessity (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 that 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 5 and at deontic logic in chapter 6. In chapter 7 we will turn to a branch of modal logic called *temporal logic* that is concerned with reasoning about time. Chapter 8 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 4, 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.

Let's return to English expressions for necessity and possibility. These often don't have the form of sentence operators. 'Must', 'might', and 'can' are (auxiliary) verbs. We can also make claims about necessity and possibility with adjectives like 'feasible', 'certain', and 'obligatory', or with adverbs like 'possibly', 'certainly', and 'inevitably'. When translating from English into $\mathfrak{L}_M$, it is often helpful to first paraphrase the English sentence in terms of '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 therefore $\neg\Diamond p$, where p represents 'you go from Auckland to Sydney by train' and the diamond represents the relevant kind of circumstantial possibility.

We often use the diamond to represent some kind of possibility, and the box to represent a corresponding kind of necessity. In such a context, you may pronounce the diamond as 'it is possible that' and the box as 'it is necessary that'. However, the same symbols are also used to represent other concepts that merely share certain structural features of possibility and necessity. To be safe, it is best to pronounce the diamond as ‘diamond’ and the box as ‘box’.

Exercise 1.4

Translate the following sentences, as well as possible, into $\mathfrak{L}_M$, assuming that the diamond expresses epistemic possibility (‘it might be that’) and the box 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 lights are on, Bob might be in his office.
(e) If the murderer escaped through the window, there must 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 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 can’t go to the station, you can’t take the train.

Special care is required when translating English sentences that contain both modal expressions and an 'if' clause. Here the surface form of English is often misleading, especially if you are used to translating 'if ... then ... ' into a material conditional $A \to B$. A good strategy is to first rephrase the English sentence so that it no longer contains any conditional expressions and then translate that paraphrase. The paraphrase, and therefore the translation, will often sound rather unlike the original sentence. 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. Duality

Consider 'Bob can't be in his office'. This might be paraphrased as 'it is not possible that Bob is in his office', which suggests that $\neg\Diamond p$ is an adequate translation (where p expresses that Bob is in his office). But the sentence might also be paraphrased as 'it is certain that Bob is not in his office', which would be translated 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' says the same as 'it is possible that not A'. Accordingly, we will assume that for any $\mathfrak{L}_M$-sentence A,

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

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

If two operators stand in the relationship expressed by (Dual1) and (Dual2), they are called **duals** of each other. In modal logic, there is a convention to use the symbols $\Box$ and $\Diamond$ only for concepts that are duals of each other.

Exercise 1.7

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 the case that . . .
(f) It will 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) tells us that $\neg\Diamond\neg p$ is equivalent to $\Box\neg\neg p$. (Here I've chosen $\neg p$ as the sentence A.) More generally, it implies that for any $\mathfrak{L}_M$-sentence A, $\neg\Diamond\neg A$ is equivalent to $\Box\neg\neg A$. We assume that logically equivalent expressions can be freely replaced by one another, even in the scope of modal operators. So we can cancel the double negation in $\Box\neg\neg A$. (Dual1) therefore implies that for any $\mathfrak{L}_M$-sentence A,

$\Box A$ is equivalent to $\neg\Diamond\neg A$.

In the same manner, (Dual2) implies that (for any $\mathfrak{L}_M$-sentence A)

$\Diamond A$ 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 1.8

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) $\neg\Box\Box p$

The concept of duality can be extended from operators to sentences and to sentence schemas. A **schema** is an $\mathfrak{L}_M$-sentence in which some sub-sentences have been replaced by upper-case schematic variables A, B, C, etc. So $\Box A$ and $\Box(\Box A \to B) \to \Box A$ are schemas.

Schemas are useful when we want to talk about all $\mathfrak{L}_M$-sentences of a certain form. The schema effectively specifies the form. Every $\mathfrak{L}_M$-sentence that results from a schema by uniformly replacing the schematic variables with object-language sentences is called an **instance** of the schema. Thus $\Box p$, $\Box\neg p$, and $\Box\Box\Box\Box(p \to p)$ are instances of $\Box A$. ("Uniformly" means that the same schematic variable is always replaced by the same object-language sentence: $p \to q$ is not an instance of $A \to A$. Different schematic variables may however be replaced by the same object-language sentence: $p \to p$ is an instance of $A \to B$.)

Exercise 1.9

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

(a) $\Box(p \to \Diamond(q \land r))$
(b) $\Box(\Diamond p \to \Diamond(\Diamond p \land p))$
(c) $\Box\Box(p \to \Diamond(p \land q))$
(d) $\Box((p \to \Diamond(p \land q)) \to \Diamond((p \to \Diamond(p \land q) \land \Diamond p)))$
(e) $\Box((A \land C) \to \Diamond((A \land C) \land (B \land C)))$

Now compare the schemas $\Box A \to A$ and $A \to \Diamond A$. (You may, if you want, informally read the first as saying that whatever is necessary is the case and the second as saying that whatever is the case is possible.) Given the duality of the box and the diamond, and our assumption that replacing logically equivalent expressions never changes the meaning of a sentence, we can show that every instance of one of them is equivalent to an instance of the other. In that sense, the two schemas are considered 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 $\Box A \to A$ is equivalent to an instance of $A \to \Diamond A$, take a simple example: $\Box p \to p$. By the truth-table for the arrow, this is equivalent to $\neg p \to \neg\Box p$. By (Dual2), $\neg\Box p$ is equivalent to $\Diamond\neg p$. So $\neg p \to \neg\Box p$ is equivalent to $\neg p \to \Diamond\neg p$. And this is an instance of $A \to \Diamond A$. The same line of reasoning obviously works for any other sentence in place of p, and a very similar line of reasoning shows the converse, that every instance of $A \to \Diamond A$ is equivalent to an instance of $\Box A \to A$.

We can generalise this observation. Any schema with the arrow $\to$ (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 1.10

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

Exercise 1.11

A proposition is *contingent* if it 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 $\Box p$.

1.4. The turnstile

So far, I have rather informally talked about validity, entailment, and equivalence. It is time to make these notions a little more precise.

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, and that an argument is logically valid if it is valid "in virtue of its logical form". What does that mean?

Consider this 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 count as 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 then we say that the premises entail the conclusion. If an argument is *logically* valid then the premises *logically* entail the conclusion. We therefore 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 all 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") to express logical consequence. The claim that $\Box(p \to q)$ and $\Box p$ together entail q, for example, could be expressed as

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

The symbol '$\models$' is not part of $\mathfrak{L}_M$, nor does $\mathfrak{L}_M$ have a comma. The comma and the turnstile belong to the **meta-language** we use to talk about the **object language** $\mathfrak{L}_M$. (The rest of our meta-language is mostly English.) We use the turnstile to express a certain relationship between $\mathfrak{L}_M$-sentences, not to construct further $\mathfrak{L}_M$-sentences.

Exercise 1.12

If we are allowed to re-interpret the non-logical expressions, do we even need to consider alternative scenarios? What do you think of this simpler alternative to definition 1.2? "$\Gamma \models A$ 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 will prove useful.

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

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

Proof. Look at the statement on the right-hand side of the 'iff'. $\Gamma \models A \to B$ says that there is no conceivable scenario in which all sentences in Γ are true while $A \to B$ is false, under any interpretation of the non-logical expressions. By the truth-table for '$\to$', $A \to B$ 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 in which all sentences in Γ are true and A is true and B is false. That's just what the statement on the left-hand side asserts. □

Observation 1.1 tells us that if we start with a claim of the form $A_1, A_2, A_3 \ldots \models B$, 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

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

we can equivalently say

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

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 $\mathfrak{L}_M$, the other to our meta-language. They also have very different meanings. In $\mathfrak{L}_M$, $p \to q$ is true iff either p is false or q is true. By contrast, $p \models q$ 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 $A \to B$ is *valid*.

Observation 1.1 tells us that we can go further from $\Box(p \to q) \models \Box p \to \Box q$ to

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

This says that $\Box(p \to q) \to (\Box p \to \Box q)$ 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.

My definitions are still somewhat imprecise. Eventually, we will want to prove various claims about entailment and validity. To this end, we will need to say more about what counts as a conceivable scenario and as an interpretation of the non-logical expressions. Let's leave this task until the next chapter.

1.5. Systems of modal logic

Whether a sentence logically entails another sentence never depends on the meaning of non-logical expressions. But it may well depend on the meaning of logical expressions. This is relevant to modal logic because the meaning of the box and the diamond – which are treated as logical expressions – varies from application to application. Sometimes the box means epistemic necessity. Sometimes it means deontic necessity. Sometimes it means circumstantial necessity. Sometimes it means yet other things.

As a result, we often can't say, once and for all, whether a certain $\mathfrak{L}_M$-sentence entails another $\mathfrak{L}_M$-sentence. Consider, for example, the inference from $\Box p$ to p. This is plausibly valid if the box expresses some kind of circumstantial necessity. It is surely invalid if the box expresses deontic necessity. 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 ($\neg p$).

To avoid confusion, we might put subscripts on the turnstile, so that $\Box p \models_c p$ says that $\Box p$ entails p in the logic of circumstantial necessity, while $\Box p \models_d p$ says (falsely) that $\Box p$ entails p in the logic of deontic necessity.

Suppose, now, that we want to fully spell one of these "logics". We want to completely specify which $\mathfrak{L}_M$-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.

The problem we face is that there are infinitely many $\mathfrak{L}_M$-sentences. To spell out our logic, we can't look at every sentence and inference one by one. We'll need to find some shortcuts.

We can begin by drawing on a consequence of observation 1.1. Above I said that we need to specify (i) which $\mathfrak{L}_M$-sentences are valid and (ii) which $\mathfrak{L}_M$-sentences are entailed by which others. Observations 1.1 implies that we can ignore part (ii) of the task. Once we have settled which sentences are valid, we have implicitly also settled which sentences entail which others. If, for example, we decide that $\Box p \to p$ is valid, we have also decided that $\Box p$ entails p.

Our task of spelling out a particular logic – a particular consequence relation – therefore reduces to the task of specifying, for every $\mathfrak{L}_M$-sentence, whether or not it is valid.

In the following chapters, we will often talk about **systems of modal logic**. Officially, a system of modal logic is simply a set of $\mathfrak{L}_M$-sentences. But these sets are of interest because they define a consequence relation, by assuming that all and only the members of the set are valid and using observation 1.1 to convert claims about validity into claims about entailment.

To make this more concrete, let's look at a particular sub-flavour of circumstantial necessity – sometimes called *historical necessity* – that was a popular object of debate in ancient Greece. Something is historically necessary if it is "settled": it is true and there is nothing we 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 somewhat 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 p is open: it could be made true; it is not settled that it is false.

Our task is to specify all $\mathfrak{L}_M$-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 $\mathfrak{L}_M$-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 $\Box p \to p$ is in. And there is nothing special here about the sentence p. Whatever is settled is true. Every instance of the schema $\Box A \to A$ is in.

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)	$\Box A \to A$
(K)	$\Box(A \to B) \to (\Box A \to \Box B)$
(4)	$\Box A \to \Box\Box A$
(5)	$\Diamond A \to \Box\Diamond A$

(Dual) corresponds to the duality principle (Dual1) from section 1.3. 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:

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

On our present interpretation, this says that if a material conditional $A \to B$ is settled, and its antecedent A is settled, then its consequent B is settled as well. To see why this is valid, let A and B be arbitrary propositions, and assume that $A \to B$ and A

are both settled. It follows that they are both true. Since $A \to B$ 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 $A \to B$ or A is false, since one of these is guaranteed to be false in any situation in which B is false. The assumption that $A \to B$ 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 open then it is settled that it is open. 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 $\Box(A \land B) \to (\Box A \land \Box B)$ 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, for example, $\Box p \to p$ is valid (in the logic of historical necessity), we can infer that $(\Box p \to p) \lor q$ is also valid, because $A \lor B$ follows from A in classical propositional logic. By adopting this rule, we effectively assume that (i) the meaning of the truth-functional connectives is given by their standard truth tables, and that (ii) every sentence is either true or false and not both. Our system of modal logic thereby becomes an **extension** of classical propositional logic.

To state the rule concisely, let $\Gamma \models_P A$ 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 $\Gamma \models_P A$ 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' and 'open') on which every sentence that is valid in the logic of historical necessity is either an instance of (T), (K), (4), (5), or (Dual), or can be generated from such instances by (CPL) and (Nec).

The system of modal logic defined by these five schemas and two 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 using 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

Which of the schemas and rules I have listed are plausible for the following interpretations of the box (with the diamond defined as the 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

Remember that a system of modal logic is just a set of $\mathfrak{L}_M$-sentences that implicitly defines a consequence relation. 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 $\Box(p \land q) \to \Box p$ 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 $(p \land q) \to p$. This is a propositional tautology, so (CPL) tells us that it is in S5. By (Nec), it follows that $\Box((p \land q) \to p)$ is in S5 as well. We also know that all instance of (K) are in S5. In particular, we have

$$\Box((p \land q) \to p) \to (\Box(p \land q) \to \Box p).$$

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

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

1.	$(p \land q) \to p$	(CPL)
2.	$\Box((p \land q) \to p)$	(1, Nec)
3.	$\Box((p \land q) \to p) \to (\Box(p \land q) \to \Box p)$	(K)
4.	$\Box(p \land q) \to \Box p$	(2, 3, CPL)

And here is a line of reasoning in the streamlined format to show that $\Box p \to \Diamond p$

is S5-valid.

1. $\Box\neg p \to \neg p$ (T)
2. $\neg\Diamond p \leftrightarrow \Box\neg p$ (Dual)
3. $\neg\Diamond p \to \neg p$ (1, 2, CPL)
4. $p \to \Diamond p$ (3, CPL)
5. $\Box p \to p$ (T)
6. $\Box p \to \Diamond p$ (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) $\Box(\Box p \to p)$
(b) $(\Box p \land \Box q) \to \Box(p \land q)$
(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 what (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, I will introduce a very different method that is much easier to use.