10 Modal logics

A deduction, Aristotle tells us, requires a conclusion which 'comes about by necessity' given some premisses. So it is no surprise that, from the very beginning, logicians have been interested in the modal notions of necessity and possibility. Modern modal logics aim, at least in the first place, to regiment reasoning about such notions. But as we will see, they can be applied much more widely.

Here's an attractive thought: *it is necessarily true that* A just if A is not only true here in the actual world but also obtains in all relevant possible worlds. Suppose we add to a logical language a symbol $\Box$, where $\Box A$ is to be read as *it is necessarily true that* A. Then, to formally model our attractive thought, we will take some objects to represent possible worlds, and say that $\Box A$ is true at 'world' w in the model just if A is true at all 'worlds' w' suitably related to w.

Compare: in §8.3(c), we described a semantic model for intuitionistic logic with the following key feature – to determine whether the conditional $A \to B$ holds in a situation k in the model, it isn't enough to know whether A holds in k and whether B holds in k; we also need to know whether A and B obtain in *other* situations k' suitably related to k. So now the idea is to use a similar relational semantics for the necessity operator, with the truth of $\Box A$ in one situation w again depending on what happens in *other* related situations w'.

In this chapter, we briefly explore this key idea by taking a look at some basic modal logics in §10.1. These and similar logics will be of interest to quite a few philosophers and also eventually to some mathematicians and computer scientists who investigate relational structures. There is, however, one rather distinctive modal logic which should be of particular interest to *anyone* beginning mathematical logic, namely so-called provability logic: so we will (even more briefly) introduce that in §10.2.

10.1 Some basic modal logics

(a) Notation first. As just proposed, we are going to add a one-place operator $\Box$ to our familiar logical languages (propositional, first-order), governed by the new syntactic rule that if A is a wff, so is $\Box A$.

Now, as we said, $\Box$ is typically going to be interpreted as some sort of necessity operator. We could also build into our languages a corresponding possibility operator $\Diamond$ (so we read $\Diamond A$ as *it is possibly true that* A). But, to keep things simple, we won't do that, since $\Diamond A$ can equally well be treated as just a definitional abbreviation for $\neg\Box\neg A$. Reflect: it is possibly true that A iff A is true at some possible world, iff it isn't the case that A is false at all possible worlds, iff it isn't the case that $\neg A$ is necessary. So the parallel between the equivalences $\Diamond/\neg\Box\neg$ and $\exists w/\neg\forall w\neg$ is not an accident!

A third modal symbol you will come across is $\prec$, for what is standardly called 'strict implication'. But again, we can treat $A \prec B$ as a definitional abbreviation, this time for $\Box(A \to B)$.

Hence, following quite common practice, we will here take $\Box$ to be the sole built-in modal operator in our languages.

(b) The story of modern modal logic begins with C. I. Lewis's 1918 book *A Survey of Symbolic Logic*. Lewis presents postulates for $\prec$, motivated by claims about the proper understanding of the idea of implication, though unfortunately his claims do seem pretty muddled.[1] Later, in C. I. Lewis and C. H. Langford's 1932 *Symbolic Logic*, there are further developments: the authors distinguish five modal logics of increasing strength, labelled *S1* to *S5*. But why multiple logics?

Let's take four schemas, and ask whether we should accept all their instances when the $\Box$ is interpreted in terms of necessary truth:

K $\Box(A \to B) \to (\Box A \to \Box B)$
T $\Box A \to A$
S4 $\Box A \to \Box\Box A$
S5 $\neg\Box A \to \Box\neg\Box A$

Well, on any understanding of the idea of necessity, if $A \to B$ and A both hold necessarily, so does B: so we can accept the principle K. And necessary truth implies plain truth: so we can accept T too. But what about the principles S4 and S5 (which are in fact distinctive of Lewis and Langford's systems *S4* and *S5*)?

It seems that different principles about repeated modalities will be acceptable depending on how exactly we interpret the necessity involved. Take a couple of examples. Suppose we interpret $\Box A$ in a mathematical context as meaning that A necessarily holds in the sense that *it is provable that* A (i.e. is provable by ordinary informal standards of proof): then arguably (i) in this case, S4 but not S5 holds. Alternatively, suppose we interpret $\Box$ as indicating analyticity in the old-fashioned philosopher's sense (where it is analytically true that A if A is true just in virtue of its conceptual content): then arguably (ii) in this case, both the S4 and S5 principles hold. But I'm certainly not going to get into the business of assessing the supposed arguments for (i) and (ii) – the issues are far too murky. And that's exactly the point to make here: the early discussions of systems of modal logic, and the supposed semantic justifications for various suggested principles, were entangled with contentious philosophical arguments. No wonder then that modal logic initially had a somewhat shady reputation!

[1] The modern reader might well suspect confusion between ideas that we now demarcate by using the distinguishing notations $\to, \vdash$ and $\vDash$.

(c) The picture radically changed some thirty years after Lewis and Langford, when Saul Kripke (in particular) developed a sharply characterized framework for giving semantic models for various modal logics.

Let's begin with the headline news about some modal *propositional* logics. In this subsection we'll describe a family of semantic models. In the next subsection we'll describe a family of deductive modal proof systems. Then the following subsection makes the Kripkean connections between the two!

So let's assume we are working in some suitable language L with the absurdity constant $\bot$ built in alongside the other usual propositional connectives, plus the unary operator $\Box$. And to define a relational semantics for such a language, we obviously need to start by introducing relational structures:

1. The basic ingredients we need are some objects W and a relation R defined over them. For the moment, think of W as a collection of 'possible worlds' and then wRw' will say that the world w' is possible relative to w (or if you like, w' is an accessible possible world, as seen from w).
2. And we will pick out an object w_0 from W to serve as the 'actual world'.

But we need an important further idea:

3. To get different flavours of relational structure (for interpreting different flavours of modal deductive system) we will want to specify different conditions S that the relation R needs to satisfy. For just one example, we might be particularly interested in relational structures where R is specified as being transitive and reflexive.

Let's say, for short, that a relational structure where the relation R satisfies the condition S is an S-structure.

Next we define the idea of a valuation of L-sentences on an S-structure. The story starts unexcitingly!

$1'$. We initially assign a value, either *true* or *false*, to each propositional letter of L with respect to each world w. Then,

$2'$. The propositional connectives behave in the now entirely familiar classical ways. For example, $A \to B$ is true at w if and only if either A is false at w or B is true at w; and so forth.

The only real novelty, as trailed at the outset, is in the treatment of the modal operator $\Box$. We stipulate

$3'$. $\Box A$ is true at a world w if and only if A is true at every world that is possible relative to w, i.e. A is true at every world w' such that wRw'.

Evidently, given ($2'$) and ($3'$), every valuation ends up assigning a value to each L-wff A at each world.

Let's say that an S-structure together with such a valuation for L-sentences is an S-model for L. Then, continuing our list of definitions, when A is an L-sentence, 4′. A is (simply) true in a given S-model for L if and only if A takes the value *true* at the actual world w_0 in the model.

Finally, and predictably, we say

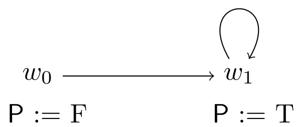
5′. A is S-valid if and only if it is true in every S-model.

So that sets up the general framework for a relational semantics for a propositional modal language. But we are now going to be interested in four different particular versions got by filling out the specification S in different ways, and so giving us four different notions of validity for propositional modal wffs:

(K) K-validity is defined in terms of K-models which allow any relation R (the specification condition S is null).
(T) T-validity is defined in terms of T-models which require the relation R to be reflexive.
(S4) *S4*-validity is defined in terms of *S4*-models which require the relation R to be reflexive and transitive.
(S5) *S5*-validity is defined in terms of *S5*-models which require the relation R to be reflexive, transitive and symmetric (i.e. R has to be an equivalence relation).

As we will soon discover, the letter-names we have chosen are indeed significant!

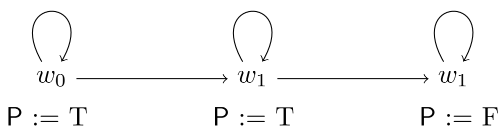
(d) Let's look at a couple of very instructive mini-examples. Take first the following two-world model, with an arrow $w \longrightarrow w'$ depicting that wRw', and with the values of P at each world as indicated:



Now, in this model, $\Box\mathsf{P}$ is true at w_0, since P is true at every world accessible from w_0, namely w_1. $\Box\mathsf{P}$ is also true at w_1, since P is again true at every world accessible from w_1, namely w_1 itself. And so $\Box\Box\mathsf{P}$ is true at w_0, since $\Box\mathsf{P}$ is true at every world accessible from w_0.

But note $\Box\mathsf{P} \to \mathsf{P}$ is *false* at w_0. So in a model like this one where the accessibility relation is not reflexive, not every instance of the schema T is true. Conversely, a moment's reflection shows that in T-models, which require that the accessibility relation is reflexive, instances of the schema T must always be true (because if $\Box A$ is true at w_0 then A is true at all accessible worlds, which will include w_0 by the reflexiveness of accessibility).

For our second example, take this three-world model:



Note, this *is* not only a K model but also a T-model, because the diagrammed accessibility relation R is reflexive; but it is not an $S4$ model since R is not transitive (we have w_0Rw_1 and w_1Rw_2 but not w_0Rw_2).

Now, in this model, $\Box\mathsf{P}$ is true at w_0 (because P is true at both the accessible-from-w_0 worlds, i.e. at w_0 and w_1). But $\Box\mathsf{P}$ is false at w_1 (because P is false at the accessible-from-w_1 world w_2). And then since $\Box\mathsf{P}$ is false at w_1 and w_1 is accessible from w_0, it follows that $\Box\Box\mathsf{P}$ is false at w_0. And hence in this model $\Box\mathsf{P} \to \Box\Box\mathsf{P}$ is false (i.e. false at w_0). Moral: the S4 principle can fail in models where the accessibility relation is not transitive.

But we can also show the reverse – in other words, in $S4$ models where the accessibility relation is transitive, the S4 principle holds. That follows because S4 can only fail in a model if the accessibility relation is non-transitive:

> Suppose something of the form $\Box A \to \Box\Box A$ *is* false in a given model, so (i) $\Box A$ is true at w_0 while (ii) $\Box\Box A$ is false at w_0. But for (ii) to hold, there must be a world w_1 such that w_0Rw_1 and (iii) $\Box A$ is false at w_1. And for (iii) to hold there must be a world w_2 such that w_1Rw_2 and (iv) A is false at w_2. But then (iv) w_2 must be 'invisible' from w_0, or else (i) couldn't hold: i.e. we can't have w_0Rw_2. In sum, for $\Box A \to \Box\Box A$ to fail we need three worlds such that w_0Rw_1, w_1Rw_2 but not w_0Rw_2 – which requires R to be non-transitive.

So our two mini-examples very nicely make the connection between a structural condition on models and the obtaining of a general modal principle such as T or S4. More about this very shortly.

(e) Since our main concern here is with the formalities, we won't delve into the arguments about which specification conditions S appropriately reflect which intuitive notions of necessity (though note that even the condition T can fail if e.g. we want to model deontic necessities – i.e. necessities of duty: since what ought to be the case may not in fact be the case!). We can leave it to the philosophers to fight things out. For now, it might be more useful to pause to summarize our semantic story in the style of our earlier account of intuitionistic semantics in §8.3(c).

So, an S-structure is a triple (w_0, W, R) where W is a set, $w_0 \in W$, and R is a relation defined over W which satisfies the conditions S. Then an S-model for a modal propositional language L is an S-structure together with a valuation relation $\Vdash$ ('makes true') between members of W and wffs of L such that

(i) $w \nVdash \bot$.
(ii) $w \Vdash \neg A$ iff $w \nVdash A$.
(iii) $w \Vdash A \wedge B$ iff $w \Vdash A$ and $w \Vdash B$.
(iv) $w \Vdash A \vee B$ iff $w \Vdash A$ or $w \Vdash B$.
(v) $w \Vdash A \to B$ iff $w \nVdash A$ or $w \Vdash B$.
(vi) $w \Vdash \Box A$ iff, for any w' such that wRw', $w' \Vdash A$.

We say that A is true in a given S-model when $w_0 \Vdash A$. As before, A is S-valid when A is true in all S-models. And for the moment the most significant conditions S on the accessibility relation R in a model are K (null), T (reflexivity), *S4* (reflexivity and transitivity), *S5* (equivalence).

Finally, I need to link up what I've just said with other presentations you'll encounter. So note that – although Kripke's original presentation did involve, as here, picking out a 'world' w_0 from W to play the role of the 'actual' world – it is clear that we can drop that step and can equivalently re-define S-validity as truth at *all* worlds in an S-model. (Why? Obviously, if A is valid on the revised definition it is valid on our original definition. While if A is not valid on the revised definition, A must be false at some world, and so it will be false on the Kripke model with that world chosen as the 'actual' world w_0.)

(f) Now let's turn to consider some proof systems for propositional modal logics. And, just because it is simplest way to do things, let's give an old-school axiomatic presentation (leaving natural deduction and tableaux versions to be explained in the recommended reading). Here then are four key systems, starting with the simplest:

(K) The modal axiomatic system K is the theory whose axioms are

(Ax i) All instances of tautologies
(Ax ii) All instances of the schema K

And whose rules of inference are

(MP) From A and $A \to B$, infer B
(Nec) If A is deducible as a theorem, infer $\Box A$

To explain briefly: (Ax i) gives us all classical tautologies and more. For read it as meaning that, given a schema for a classical tautology, the result of systematically substituting *any* wffs of our modal propositional language for schematic letters – even substituting modalized wffs – will be an axiom of K. So, for example, $(A \wedge B) \to A$ is a schema for a classical tautology. Hence the result of substituting $\Box\mathsf{P}$ for A and $\Box\mathsf{Q}$ for B, giving us $(\Box\mathsf{P} \wedge \Box\mathsf{Q}) \to \Box\mathsf{P}$, is an axiom of K. Such instances of tautologies are still, surely, logical truths.

We've already said that instances of (Ax ii) look good on any suitable reading of the box. And our old friend the modus ponens rule (MP) is uncontentious.

Which leaves the necessitation rule (Nec). This is to be very sharply distinguished from what would evidently be the quite unacceptable axiom schema $A \to \Box A$: obviously, A can be true without being necessarily true. However, the idea justifying (Neg) is that if A is actually a logical theorem – i.e. is deducible from logical principles alone – then it will indeed be necessary (on almost any sensible understanding of 'necessary'). Here's an example of the rule (Nec) in use in a K-proof:

1.	$((\mathsf{P} \wedge \mathsf{Q}) \to \mathsf{P})$	Axiom, by (Ax i)
2.	$\Box((\mathsf{P} \wedge \mathsf{Q}) \to \mathsf{P})$	By (Nec), since 1 is a theorem
3.	$\Box(((\mathsf{P} \wedge \mathsf{Q}) \to \mathsf{P}) \to (\Box(\mathsf{P} \wedge \mathsf{Q}) \to \Box\mathsf{P}))$	Axiom, by (Ax ii)
4.	$(\Box(\mathsf{P} \wedge \mathsf{Q}) \to \Box\mathsf{P})$	From 2 and 3 by (MP)
5.	$\Box(\Box(\mathsf{P} \wedge \mathsf{Q}) \to \Box\mathsf{P})$	By (Nec), since 4 is a theorem

In sum, then, all the theorems of the weak system K – i.e. all the wffs deducible from axioms alone – should be logical truths on (almost all) readings of $\Box$ read as a kind of necessity.

And now here are three nested ways of strengthening the system K:

(T) T is the axiomatic system K augmented with all instances of the schema T as axioms.
(S4) *S4* is T augmented with all instances of the schema S4 as axioms.
(S5) *S5* is *S4* augmented with all instances of the schema S5 as axioms.

(g) So now at last for the big reveal – except of course I've entirely sploit any element of surprise by the parallel labelling of the flavours of modal semantics and the flavours of axiomatic proof system!

What Kripke famously showed is the following lovely result:

> Whether S is K, T, *S4*, *S5*, a wff A is an S-theorem if and only if it is S-valid.

In short, we have soundness and completeness theorems for our proof systems. And there are some nice immediate implications. Searching for an appropriate countermodel which shows that a wff is not S-valid is a finite business, so it is decidable what's S-valid – and hence it is decidable what's an S-theorem.[2]

These soundness and completeness results are not mathematically very difficult. Perhaps Kripke's real achievement was the prior one in developing the general semantic framework and in finding the required simple proof systems – some of them different from any of the systems proposed by Lewis and Langford – thereby making his very elegant result possible.

(h) And now, with the apparatus of relational semantics available, the floodgates open! After all, the objects in a S-model don't have to represent 'possible worlds' (whatever they are conceived to be); they can stand in for any points in a relational structure. So perhaps they could represent states of knowledge, points of a time series, positions in a game, states in the execution of a program, levels in a hierarchy . . . with different classes of accessibly relations appropriate for different cases and so with different deductive systems to match. The resulting applications of propositional modal logics are indeed very many and various: some of the suggested readings will tell you much more.

(i) What about quantified modal logic, where we add the modal operator $\Box$ to a first-order language? Why might we be interested in them?

Well, philosophers make play with questions like this: Does it make sense to suppose the very same objects can appear in the domains of different possible

[2]Suppose we define in the now obvious ways (i) the idea of a conclusion being an S-valid consequence of some finite number of premisses, and (ii) the idea of that conclusion being deducible in system S from those premisses. Then again we have soundness and weak completeness proofs linking valid consequences with deductions, and we have corresponding decidability results too. We won't worry however about strong completeness, which does indeed fail for some modal logics, e.g. for GL which we meet in the next section¿

worlds? If it does, do all possible worlds contain the same objects (perhaps some of them actualized, some not)? Does a proper name (formally a constant term) denote the same thing at any possible world at which it denotes at all? Are atomic identity statements, if true at all, necessarily true? Questions of this stripe pile up, and they motivate different ways of tweaking quantified modal logic in formally modelling and so clarifying the philosophical ideas.

However, the resulting logics don't seem to be of particular interest to non-philosophers; the wider logical community has (as yet) been *much* more interested in propositional modal logics.

Still, the beginnings of the technical story about first-order modal logics are pretty accessible. And the suggested readings will enable you to get some headline news about different proof systems and their formal semantics, without getting too entangled in unwanted philosophical debates!

10.2 Provability logic: the very idea

Propositional modal logics, to repeat, have a wide range of applications. But there is one that stands out as being of pre-eminent relevance to anyone beginning mathematical logic. And that is provability logic.

(a) Let's start with some reminders of what you should already know from tackling Gödel's incompleteness theorems (see §6.4). So take a theory in which we can do enough arithmetic – to fix on an example, take first-order Peano Arithmetic. Choose a sensible system of Gödel-numbering. Then you can construct a relational predicate in the language of arithmetic – one which we can abbreviate $\mathsf{Prf}(\mathsf{x},\mathsf{y})$ – that nicely[3] represents the relation which obtains between two numbers x, y, when x is the Gödel number of a PA proof of the sentence with Gödel number y. Now define $\mathsf{Prov}(\mathsf{y})$ to be the expression $\exists \mathsf{x}\mathsf{Prf}(\mathsf{x},\mathsf{y})$. Then $\mathsf{Prov}(\mathsf{y})$ represents the property that a number y has if it numbers a theorem of PA – so Prov is naturally called a *provability predicate*.

If A is wff of arithmetic, let $\ulcorner A \urcorner$ be shorthand for A's Gödel-number, and let $\overline{\ulcorner A \urcorner}$ be shorthand for the formal numeral for $\ulcorner A \urcorner$. Then, given our definitions, $\mathsf{Prov}(\overline{\ulcorner A \urcorner})$ says that A is provable in PA.

Now we introduce yet another bit of shorthand: let's use $\boxdot A$ as a simple abbreviation for $\mathsf{Prov}(\overline{\ulcorner A \urcorner})$.[4] With some effort, we can then show that PA proves (unpacked versions of) all instances of the following familiar-looking schemas

K· $\boxdot(A \to B) \to (\boxdot A \to \boxdot B)$
S4· $\boxdot A \to \boxdot \boxdot A$

And moreover we have an analogue of the modal Necessitation rule:

(Nec·) If A is deducible as a PA theorem, then so is $\boxdot A$.

---

[3]'Nicely' waves a hand at some details which are important but which we won't need to delay over here!

[4]I've dotted the box here – not the usual notation – for clarity's sake!

That package of facts about PA is standardly reported by saying that the theory satisfies the so-called HBL derivability conditions. And appealing to these facts together with the First Incompleteness Theorem, it is then easy to derive the Second Theorem that PA cannot prove $\neg \boxdot \bot$ (i.e. can't prove that $\bot$ isn't provable, i.e. can't prove that PA is consistent).[5]

(b) The obvious next question might well seem to be: what *other* modal principles/rules should our dotted-box-as-a-provability-predicate obey, in addition to the dotted principles $\mathsf{K}\cdot$ and $\mathsf{S4}\cdot$, and the rule (Nec$\cdot$)? What is its appropriate modal logic?

But hold on! We are getting ahead of ourselves, because we so far only have the illusion of modal formulas here. The box as just defined simply doesn't have the right grammar to be a modal operator. Look at it this way. In a proper modal language, the operator $\Box$ is applied to a wff A to give a complex wff $\Box A$ in which A appears as a subformula. But in our newly defined usage where $\boxdot A$ is short for $\mathsf{Prov}(\overline{\ulcorner A \urcorner})$, the formula A doesn't appear at all – what fills the appropriate slot(s) in the predicate Prov is a *numeral* (the numeral for the number which happens to code the formula A).

In short, the surface form of our notation $\boxdot A$ is entirely misleading as to its logical form. Which is why the logically pernickety might not be very happy with it.

However, it remains the case that our abbreviatory notation is highly suggestive. And what it suggests is starting with a kosher modal propositional language of the now familiar kind, where the box is genuinely a unary operator applied to wffs. And then we consider *arithmetical interpretations* which map sentences A of our modal language to corresponding sentences A^* of PA, interpretations which have the following shape:

i. An interpretative map sends an atomic letter A of our modal language to some arithmetical sentence A^*, any you like.
ii. The map then respects the propositional connectives: for example, it sends conjunctions in the modal language to conjunctions in the arithmetic language, so $(A \wedge B)^*$ is $(A^* \wedge B^*)$; it sends the absurdity constant to the absurdity constant, i.e. $\bot^*$ is $\bot$; and so on.
iii. The map sends the modal sentence $\Box A$ to $\boxdot A^*$, i.e. to $\mathsf{Prov}(\overline{\ulcorner A^* \urcorner})$.

There is now no notational jiggery pokery; we have a respectable modal language on the one side, and various interpretative mappings from its sentences into a regular arithmetical language on the other side.

And *now* we can ask a cogent versions of the misplaced question we wanted to ask before. In particular, we can ask: what are the modal sentences which are such that, on *any* interpretative mapping into PA, their translations are arithmetical theorems? What, for short, is the correct modal logic for the $\Box$ interpreted *this* way as tracking formal provability in PA?

[5]For more details, if this is new to you, see for example Chapter 33 of my *An Introduction to Gödel's Theorems* (downloadable from logicmatters.net/igt).

(c) Here's a reminder of another result we can get from the HBL conditions, namely Löb's Theorem.

Momentarily using again our now rather deprecated dotted-box-as-abbreviation notation, this rather surprising theorem says:

If PA proves $\boxdot A \to A$, then it proves A.[6]

We will presumably want to reflect this theorem in a logic for the genuinely modal $\Box$ operator interpreted as arithmetical provability: a natural move, then, is to build into our modal logic the rule that, if $\Box A \to A$ is deducible as a theorem, then we can infer A.

So this putting this thought together with our previous remarks, let's consider the following modal logic – the 'G' in its name is for Gödel who made some prescient remarks, and the 'L' is for Löb:

(GL) The modal axiomatic system *GL* is the theory whose axioms are

- (Ax i) All instances of tautologies
- (Ax ii) All instances of the schema K: $\Box(A \to B) \to (\Box A \to \Box B)$
- (Ax iii) All instances of the schema S4: $\Box A \to \Box\Box A$

And whose rules of inference are

- (MP) From A and $A \to B$, infer B
- (Nec) If A is deducible as a theorem, infer $\Box A$
- (Löb) If $\Box A \to A$ is deducible as a theorem, infer A.

You can immediately see, by the way, that we *don't* also want to add all instances of the T-schema $\Box A \to A$ to this modal logic. For a start, doing that would make $\Box\bot \to \bot$ a theorem and hence $\neg\Box\bot$ would be a theorem. But that can't correspond on arithmetic interpretation to a theorem of PA, since we know that PA can't prove $\neg \boxdot \bot$.

And there's worse: leaving aside the desired interpretation of this logic, if we add all instances of $\Box A \to A$ as axioms, then in the presence of the rule (Löb), we can derive any A, and the logic is inconsistent.

Now, given our motivational remarks in defining *GL*, it won't be a surprise to learn that it is indeed *sound* on the provability interpretation. Once we have done the (non-trivial!) background work required for showing that the HBL derivability conditions and hence Löb's theorem hold in PA, it is quite easy to go on to establish that, on every interpretation of the modal language into the language of arithmetic, every theorem of *GL* is a theorem of PA.

And (with more decidedly non-trivial work due to Robert Solovay) it can also be shown that *GL* is *complete* on the provability interpretation. In other words, if a modal sentence is such that every arithmetic interpretation of it is a PA theorem, then that sentence is a theorem of the modal logic *GL*.

Which is all very pleasingly neat!

[6]See Chapter 34 of *An Introduction to Gödel's Theorems.*

(d) We should pause to note that there is another way of presenting this provability logic.

Suppose we drop the Löb inference rule from GL, and replacethe instances of the S4 schema as axioms with instances of the Löb-like schema

L $\Box(\Box A \to A) \to \Box A$

It is then quite easy to see that this results in a modal logic with exactly the same theorems (because GL in our original formulation implies all instances of L; and conversely we can show that all instances of S4 can be derived in the new formulation, for which the Löb rule is also a derived rule of inference). Hence either formulation gives us the provability logic for PA.

(e) Now, we've so far been working with arithmetic interpretations of our modal wffs. But we can also give a more abstract Kripke-style relational semantics for GL (it is a nice question, though, whether this 'semantics' has much to do with meaning!). We start by defining a GL-model in the usual sort of way as comprising a valuation with respect to some worlds W with a relation R defined over them, where R satisfies ...

Well, what conditions do we in fact need to place on R so that GL-theorems match with the GL-validities (the truths that hold at every world, on every GL-model)? Clearly, we *mustn't* require R to be reflexive – or else all instances of the T-schema would come out GL-valid, and we don't want *that*. Equally clearly, we *must* require R to be transitive – or else instances of the S4-schema could fail to be GL-valid. But we need more: what further condition on R is required to make all the instances of the L-schema come out valid?

It turns out that what is needed is that there is no infinite chain of R-related worlds $w_0, w_1, w_2, w_3, \ldots$ such that $w_0 R w_1 R w_2 R w_3 \ldots$ (and that condition ensures that R is irreflexive, for otherwise we would have some infinite chain $wRwRwRw\ldots$). Call that the finite chain condition. Then define a GL-model as one where the accessibility relation R is transitive and satisfies the finite chain condition. Then a modal sentence is GL theorem if and only if it is GL-valid (true in all worlds in all GL-models).

This new soundness and completeness theorem has a lovely upshot. As with the other modal logics we've met, there is a systematic way of testing for GL-validity (by systematically searching for countermodels). So it is decidable what's a GL theorem.

(f) That last result, together with the fact that GL is sound and complete for arithmetical interpretations into theorems of PA, shows something rather remarkable. Although PA as a whole is an undecidable theory, there is a very interesting *part* of that theory – roughly, what it can say by applying propositional logic and its provability predicate to arithmetical wffs – which *is* decidable.

For example, consider this question: for any arithmetical sentence A, does PA know – i.e. can it prove? – that if A is provably equivalent to the claim it isn't provable, that A is provably equivalent to saying that PA is consistent? In symbols, using the dotted-box-as-abbreviation, can PA prove

$$\boxdot(A \leftrightarrow \neg \boxdot A) \to \boxdot(A \leftrightarrow \neg \boxdot \bot)$$

Well it can so long as the corresponding modal wff

$$\Box(\mathsf{P} \leftrightarrow \neg\Box\mathsf{P}) \to \Box(\mathsf{P} \leftrightarrow \neg\Box\bot)$$

is a *GL* theorem – and that's decidable (in fact, it *is* a theorem).

This way, we easily get to know a lot more about what PA can prove about it can prove. And this is just one example of the kind of payoff we get from applying modal logic to questions of provability in arithmetics. Hence the considerable interest of provability logic.

10.3 First readings on modal logic

(a) There is, as so often, a good entry in that wondrous resource the Stanford encyclopaedia, one which should provide more very helpful orientation:

1. James W. Garson, 'Modal logic', *Stanford Encyclopedia of Philosophy*: read §§1–11 and 15. Available at tinyurl.com/sep-modal.

Now, because of its interest, modal logic is often taught to philosophers without much logical background, and so there are a number of introductions written primarily for them. A good example is the very accessible

2. Rod Girle, *Modal Logics and Philosophy* (Acumen 2000; 2nd edn. 2009). Part I of this book provides a particularly lucid introduction, which in 136 pages explains the basic syntax and relational semantics, covering both trees (tableaux) and natural deduction for some propositional modal logics, and extends as far as the beginnings of quantified modal logic.

Philosophers may well very want to go on to read Part II of this book, on applications of modal logic.

There is an equally lucid account in an extraordinarily useful book by Graham Priest. I'll highlight this not only because it is brisker on modal logics, but because we also get an account of intuitionistic logic in the same tableaux framework:

3. Graham Priest, *An Introduction to Non-Classical Logic** (CUP, much expanded 2nd edition 2008). This treats a whole range of logics systematically, concentrating on semantic ideas, and using a tableaux approach. Chs. 1 and 12 provide quick revision tutorials on tableaux for classical propositional and predicate logic. Then Chs. 2 and 3 give the basics on propositional modal logics. You can then either fill in more about modal logics in Ch 4 or skip to Ch. 6 on propositional intuitionistic logic. Then Chs 14 and 15 introduce the basics on quantified modal logics. You can then fill in more about quantified modal logics in Chs 16–18 or can then

skip to Ch. 20 on quantified intuitionistic logic.

This whole book – which we will revisit in our next chapter – is a terrific achievement and enviably clear and well-organized.

Then, going half-a-step up in sophistication, though still starting from scratch, we find another excellent book (elegantly done in a way which might appeal more to mathematicians):

4. Melvin Fitting and Richard L. Mendelsohn, *First-Order Modal Logic* (Kluwer 1998). This gives both tableaux and axiomatic systems for various modal logics, in an approachable style and with lucid discussions of options at various choice points. Despite its more mathematical flavour, the book still includes some interesting discussions of the conceptual motivations for different modal logics.

   Read the first half of this book to get a compact but sufficient introduction to propositional modal logics, and also the initial headlines about quantified modal logics. Philosophers will then want to read on.

(b) And now, as noted in the previous sections of this chapter, the path onwards forks. Some philosophers will want to know more about propositional modal logics and the variety of their applications (e.g. to tense logic, epistemic logic, etc.); other philosophers will be more concerned with quantified modal logics in their role of disciplining thoughts about the metaphysics of modality; while those interested in e.g. theoretical computer science will be interested in modal logics as a framework for theorizing about relational structures. I'll say a little about some relevant readings on these topics in the next section.

But here let's move on to the application of modal logic likely to be of most interest to mathematical logicians. This is nicely introduced in:

5. Rineke Verbrugge, 'Provability Logic' §§1–4 and perhaps §6, *Stanford Encyclopedia of Philosophy*. Available at tinyurl.com/prov-logic.

Or you could dive straight into the very first published book on our topic, which I think still makes for the most attractive entry-point:

6. George Boolos, *The Unprovability of Consistency* (CUP, 1979), particularly Chs 1–12. This is a famous modern classic, yet very approachable. Boolos has an admirably lucid and engaging presentational style (and the book can be read surprisingly quickly if you are happy to initially skip some of the longer proofs).

10.4 Alternative and further readings

(a) *Other introductory readings for philosophers* The first part of Theodore Sider's *Logic for Philosophy** (OUP, 2010) is poor as an introduction to FOL.

However, the second part, which is entirely devoted to modal logic and related topics like Kripke semantics for intuitionistic logic, is very much better, and philosophers could indeed find it useful. For example, the chapters on quantified modal logic (and some of the conceptual issues they raise) are brief, lucid and approachable.

Sider is, however, closely following a particularly clear old classic by G. E. Hughes and M. J. Cresswell *A New Introduction to Modal Logic* (Routledge, 1996, updating their much earlier book). This can still be recommended and may suit some readers, though it does take a rather old-school approach.

If your starting point has been Priest's book or Fitting/Mendelson, then you might want at some point to supplement these by looking at a treatment of natural deduction proof systems for modal logics. One option is to dip into Tony Roy's long article 'Natural Derivations for Priest', in which he provides ND logics corresponding to the propositional modal logics presented in tree form in Priest's book: accessible at tinyurl.com/roy-modal. But a smoother introduction to ND modal systems is provided by Chapter 5 of Girle, or by my main alternative recommendation for philosophers, namely

7. James W. Garson, *Modal Logic for Philosophers** (CUP, 2006; 2nd end. 2014). This again is intended as a gentle introductory book: it deals with both ND and semantic tableaux (trees), and covers quantified modal logic. It is indeed pretty accessible, though on balance I probably do still prefer Girle and Priest at this level.

(b) *Modal logics for philosophical applications* If you are interested in applications of propositional modal logics to tense logic, epistemic logic, deontic logic, etc. then there is a masterly compendious survey:

8. Lloyd Humberstone, *Philosophical Applications of Modal Logic** (College Publications, 2015). This starts with a book-within-a-book, an advanced 176 page introduction to propositional modal logics. And then there are extended discussions of a wide range of applications of these logics that have been made by philosophers.

If your interests instead lean to modal metaphysics, then – once upon a time – a discussion of quantified modal logic at the level of Fitting/Mendelsohn would have probably sufficed. And for more on first-order quantifed modal logics, see

9. James W. Garson, 'Quantification in Modal Logic' in *Handbook of Philosophical Logic*, edited by Dov M. Gabbay and F. Guenther (Springer, 2nd edition 2001).

However, Timothy Williamson's notable book *Modal Logic as Metaphysics* (OUP, 2013) calls on rather more, including e.g. second-order modal logics. However, there doesn't seem to be general guide/survey of higher-order modal logics at the right sort of level, with the right sort of coverage to recommend here. There *is* a text by Nino B. Cocchiarella and Max A. Freund, *Modal Logic: An Introduction* *to its Syntax and Semantics* (OUP, 2008), whose blurb announces that “a variety of modal logics at the sentential, first-order, and second-order levels are developed with clarity, precision and philosophical insight”. However, the treatments in this book are relentlessly and rebarbatively formal. In its last two chapters, the book does cover second-order modal logic: but the highly unfriendly mode of presentation will probably put the discussion out of reach of most philosophers who might be interested. You have been warned.

(c) *Provability logic again* For another introductory overview, with indications of more recent work on the topic, you could look at

10. Sergei Artemov, ‘Modal logic in mathematics’ §§1–5, in *The Handbook of Modal Logic*, edited by P. Blackburn et al. (Elsevier, 2005).

But the two obvious sources to mention here are a couple of classic books:

11. Craig Smoryński, *Self-Reference and Modal Logic* (Springer-Verlag, 1985). This is a lovely alternative or accompaniment to Boolos’s 1979 book. Well, not lovely to look at, as it oddly printed in extremely small type emulating an electric typewriter, which doesn’t make for comfortable reading. But the content is extremely lucidly and elegantly presented, with a lot of helpful explanatory/motivating chat alongside the more formal work. Also highly recommended.
12. George Boolos, *The Logic of Provability* (CUP, 1993). This is a significantly expanded and updated version of his earlier book. And indeed, you could read the first half of this instead, though I do retain a fondness for the shorter version. The main occasion for the update is the presentation of proofs of major results about quantified provability logic which were discovered after Boolos wrote his first book: but these results are really more than you need in a first encounter with provability logic.

(d) *Finally, three more technical books* In order of publication, here are three more advanced and rather more challenging texts I can suggest to sufficiently interested readers:

13. Sally Popkorn, *First Steps in Modal Logic* (CUP, 1994). The author is, at least in this possible world, identical with the late mathematician Harold Simmons. This book, which entirely on propositional modal logics, is written for computer scientists. The Introduction rather boldly says ‘There are few books on this subject and even fewer books worth looking at. None of these give an acceptable mathematically correct account of the subject. This book is a first attempt to fill that gap.’ This considerably oversells the case: but the result is illuminating and readable.
14. Alexander Chagrov and Michael Zakharyaschev *Modal Logic* (OUP, 1997). This is a volume in the Oxford Logic Guides series and again concentrates

on propositional modal logics. Definitely written for the more mathematically minded reader, it tackles things in an unusual order, starting with an extended discussion of intuitionistic logic, and is good but rather demanding.

15. Patrick Blackburn, Maarten de Ricke and Yde Venema's *Modal Logic* (CUP, 2001). This is one of the Cambridge Tracts in Theoretical Computer Science: but don't let that provenance put you off! This is an accessibly and agreeably written text on propositional modal logic – certainly compared with the previous two books in this group – with a lot of signposting to the reader of possible routes through the book, and with interesting historical notes. I think it works pretty well, and will also give philosophers an idea about how non-philosophers can make use of propositional modal logic.

10.5 Finally, a very little history

Especially for philosophers, it is very well worth getting to know a little about how mainstream modern modal logic emerged from the to-and-fro between philosophical debate and technical developments. So do read e.g. one of

16. Roberta Ballarin, 'Modern Origins of Modal Logic', *Stanford Encyclopedia of Philosophy*. Available at tinyurl.com/mod-orig.
17. Sten Lindström and Krister Segerberg, 'Modal Logic and Philosophy' §1, in *The Handbook of Modal Logic*, edited by P. Blackburn et al. (Elsevier, 2005).