5

Axioms, proofs, and completeness

5.1 Describing validities by proofs

Universal validity of a formula φ was defined somewhat abstractly as truth of φ at each world in each model. How can we describe the form of these validities more concretely? After all, logic is also about valid arguments, and premises $\varphi_1, \ldots \varphi_k$ imply conclusion ψ iff the implication $(\varphi_1 \wedge \cdots \wedge \varphi_k) \to \psi$ is a valid formula. One concrete method in logic is this: give a *proof system*, that is, a concrete set of initial principles and derivation rules that produce only valid principles (this property is called *soundness*) - and hopefully also, all of the valid principles (the famous property of *completeness*). Logical proof systems exist in many different formats: our "sequent calculus" in Chapter 4 was an example. Not all logics have complete proof systems, but there is no reason not to try in the case of modal logic.

5.2 A short-cut through first-order logic?

But perhaps we do not have to try at all? One quick, but sneaky route is as follows. Using the method of Chapter 7 (but in your heart, you already know how to do this) *translate* modal formulas into first-order ones, and then use any complete proof system that you have learnt for the latter system to derive the (translated) modal validities.

Example (Modal distribution law)**.** Instead of proving the semantically valid modal distribution law $\Box(p \to q) \to (\Box p \to \Box q)$, one can easily derive its first-order translation (note how modal boxes become successive bounded universal quantifiers here) $\forall x(\forall y(Rxy \to (Py \to Qy)) \to (\forall y(Rxy \to Py) \to \forall y(Rxy \to Qy)))$ – using only standard axioms and rules of first-order logic.

But this does not give much insight into the peculiarities of modal reasoning, which is, amongst other things, done in *variable-free no-*

tation. Moreover, a first-order proof for a translated modal formula might contain "junk": intermediate formulas that have no modal counterparts, which offends our sense of purity.[33] Therefore, we also want to find more intrinsically modal proof systems. Nevertheless, a comparison with, say, axioms for *FOL* is useful. We shall appreciate better what we need, and what not. For concreteness, Herbert Enderton's famous textbook *A Mathematical Introduction to Logic* (Enderton, 1971), used by many generations of Stanford students, has the following set:

(a) all tautologies of classical logic,
(b) distribution: $\forall x(\varphi \to \psi) \to (\forall x\varphi \to \forall x\psi)$,
(c) universal instantiation: $\forall x\varphi \to [t/x]\varphi$, provided that t is freely substitutable for x in φ,
(d) vacuous universal generalization: $\varphi \to \forall x\varphi$, provided that x is not free in φ,
(e) a definition of $\exists x\varphi$ as $\neg\forall x\neg\varphi$,[34] and
(f) the rule of Modus Ponens: "from φ and $\varphi \to \psi$, conclude ψ.

Here, each axiom can come with any finite prefix of universal quantifiers. This special feature provides the effect of the rule of

(g) Universal Generalization: "if φ is provable, then so is $\forall x\varphi$".

The syntactic provisos on Axioms (c) and (d) are a common source of errors, and they reflect the fact that the first-order language is all about variable dependency and variable handling.

Theorem 7. A first-order formula is valid iff it is provable using the Enderton axioms.

We will present a variable-free proof system for the modally valid formulas. Even so, many systems for *automated deduction* do use translation into first-order logic, since computational techniques have been highly optimized for the latter widely used system – and a user need not care so much what happens "under the hood" of the computer.

5.3 The minimal modal logic

Our basic modal proof system is like part of Enderton's complete set, but without syntax worries:

Definition 5.3.1 (Minimal modal logic)**.** The minimal modal logic K is the proof system with the following principles:

(a) all tautologies from propositional logic,

---

[33]In fact, this junk is almost bound to occur in a proof for modal distribution.
[34]Or one can make both quantifiers primitives, with an axiom $\exists x\varphi \leftrightarrow \neg\forall x\neg\varphi$.

(b) modal distribution $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$,
(c) a definition of $\Diamond\varphi$ as $\neg\Box\neg\varphi$,
(d) the rule of Modus Ponens,
(e) and a rule of Necessitation: "if φ is provable, then so is $\Box\varphi$".

Proofs are finite sequences of formulas, each of them either (i) an instance of an axiom, or (ii) the result of applying a derivation rule to preceding formulas. A formula φ is provable: written as $\vdash \varphi$, if there is a proof ending in φ. If we want to indicate the specific modal logic we are using, we write it as a subscript: for instance, $\vdash_K \varphi$.

Our variable-free modal notation has no laws like the above first-order (c) and (d). These do appear, in a sense, in stronger systems. If you wish, the axiom $\Box\varphi \to \varphi$ of the stronger modal logic T is an instance of universal instantiation.[35] Likewise, an *S5*-axiom like $\Diamond p \to \Box\Diamond p$, valid on models where the accessibility relation holds between all worlds, is really the vacuous generalization $\exists x Px \to \forall x \exists x Px$.

5.4 The art of formal proof

Finding formal proofs is a skill that can be drilled into students, and though it has few practical applications, it has a certain unworldly beauty. We will not emphasize this drill here, but the student will do well to study a few derivations in detail, and see the bag of useful tricks that goes into them. Roughly speaking, proofs in the minimal modal logic often have a propositional core, which is then "lifted" to the modal setting. Many textbooks provide examples: say, the *Manual of Intensional Logic* (van Benthem, 1988a) has a few annotated ones.

Learning formal proof is a matter of practice. You build up a library of useful sub-routines, you learn to recognize formal patterns (in fact, logic courses have been used as a laboratory for a variety of cognitive psychology experiments) and soon you are airborne.[36]

Example (Distribution rules)**.**

(a) If $\varphi \to \psi$ is provable, then so is $\Box\varphi \to \Box\psi$:

1)	$\varphi \to \psi$	provable by assumption
2)	$\Box(\varphi \to \psi)$	Necessitation rule on 1
3)	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	modal distribution axiom
4)	$\Box\varphi \to \Box\psi$	Modus Ponens on 2, 3

[35] But this often-cited analogy is not quite right, if you think it through: why?

[36] If all else fails, you can opportunistically seek an informal semantic argument for inspiration, and hide the idea in formal steps later.

(b) If $\varphi \to \psi$ is provable, then so is $\Diamond\varphi \to \Diamond\psi$:

1)	$\varphi \to \psi$	provable by assumption
2)	$\neg\psi \to \neg\varphi$	propositional logic, 1
3)	$\Box\neg\psi \to \Box\neg\varphi$	by the subroutine (a)
4)	$\neg\Box\neg\varphi \to \neg\Box\neg\psi$	propositional logic, 3
5)	$\Diamond\varphi \to \Diamond\psi$	definition of $\Diamond$

Related useful observations about modal provability include the widely used principle of

Replacement by Provable Equivalents:
if $\vdash \alpha \leftrightarrow \beta$ then $\vdash \varphi[\alpha] \leftrightarrow \varphi[\beta]$.[37]

Next, as for proving real theorems, it often helps to start at the end, and first reformulate what we are after. This is of course, standard heuristics: reformulate the result to be proved in a *top-down manner*, until you see *bottom-up* which available principles will yield it:

Example (An actual theorem of K)**.** Using these observations, we show that $\vdash_K (\Diamond\varphi \wedge \Box(\varphi \to \psi)) \to \Diamond\psi$:

1) by propositional logic, it suffices to prove the equivalent $\Box(\varphi \to \psi) \to (\Diamond\varphi \to \Diamond\psi)$, which is again equivalent to
2) $\Box(\varphi \to \psi) \to (\neg\Box\neg\varphi \to \neg\Box\neg\psi)$, which is equivalent to
3) $\Box(\varphi \to \psi) \to (\Box\neg\psi \to \Box\neg\varphi)$
4) Now we recognize a propositional core tautology that we can use: $(\varphi \to \psi) \to (\neg\psi \to \neg\varphi)$, and apply our distribution rule to it:
5) $\Box(\varphi \to \psi) \to \Box(\neg\psi \to \neg\varphi)$, and combining this with a distribution axiom to obtain:
6) $\Box(\varphi \to \psi) \to (\Box\neg\psi \to \Box\neg\varphi)$, we get the desired conclusion.

The very typographical cut-and-pastes that you will do in typing up these proofs show the workings of (i) proof structure, (ii) pattern recognition, (iii) modularity, and (iv) sub-routines!

Other well-known theorems of the minimal logic K are principles such as the distribution of $\Box$ over $\wedge$:

$$\Box(\varphi \wedge \psi) \to (\Box\varphi \wedge \Box\psi)$$

and its diamond counterpart $\Diamond(\varphi \vee \psi) \to (\Diamond\varphi \vee \Diamond\psi)$. The latter can also be derived from the former by a more general system property of *Duality*, just as in classical logic. Finally, another simple way of "seeing" K-theorems is through analogies with first-order logic.

[37] Here $\varphi[\alpha]$ is a formula containing one of more occurrences of the sub-formula α, and $\varphi[\beta]$ results from $\varphi[\alpha]$ by replacing all of these by occurrences of β.

5.5 Proofs in other modal logics

Stronger modal logics increase deductive power by adding further axiom schemata to the minimal logic K. Then the flavour of finding derivations may change, as you develop a feeling for what the additional syntactic power gives you. Here are some well-known examples:

Example (T, $S4$ and $S5$)**.** The modal logic T arises from K by adding the axiom schema of Veridicality $\Box\varphi \rightarrow \varphi$. The logic $S4$ adds the schema $\Box\varphi \rightarrow \Box\Box\varphi$ to T, which for knowledge is called Positive Introspection. Finally, the logic $S5$ adds the schema $\Diamond\varphi \rightarrow \Box\Diamond\varphi$ to $S4$.

We will discuss these logics later on, but here is one illustration:

Fact. The following principle is provable in $S4$: $\Box\Diamond\Box\Diamond\varphi \leftrightarrow \Box\Diamond\varphi$

Proof. The main steps are these. *From left to right.* (a) First, prove that the formula $\Box\alpha \rightarrow \Diamond\alpha$ is provable for all formulas α (this is easy to do even in the modal logic T), so that we have $\Box\Diamond\varphi \rightarrow \Diamond\Diamond\varphi$. Next (b) apply earlier sub-routines to get $\Box\Diamond\Box\Diamond\varphi \rightarrow \Box\Diamond\Diamond\Diamond\varphi$. Then (c) derive $\Diamond\Diamond\Diamond\varphi \rightarrow \Diamond\varphi$ in $S4$, and apply an earlier sub-routine to get $\Box\Diamond\Box\Diamond\varphi \rightarrow \Box\Diamond\varphi$. *From right to left.* "Blow up" the initial box in $\Box\Diamond\varphi$ to three boxes, using the $S4$-axiom. Then replace the second of these by a diamond, using the same principles as before. $\Box$

You can see more concretely what is going on by a "picturesque" *semantic argument* for the validity of $\Box\Diamond\Box\Diamond\varphi \leftrightarrow \Box\Diamond\varphi$ on reflexive and transitive models for $S4$. There is even a kind of heuristic translation between the semantic argument and a syntactic proof. But a semantic argument may give you more. If you analyse things well in your pictures, you will see that reflexivity is not needed.

Perhaps you can also find shorter, or otherwise different proofs for the above semantic validities. Theorems in a logical system may well have more than one non-equivalent proof!

5.6 The science of proof

On the practical side, computers can search for proofs in much faster, mechanical ways. Existing theorem provers for modal logic use various techniques.[38] Usually, these are not in the above axiomatic style, but they use a variety of other methods: (a) translation into first-order logic plus "resolution" methods, (b) the above "semantic tableaux", which are formal versions of semantic decision procedures, or (c) specially optimized modal calculi. On the theoretical side, there is a field of

[38] You can look up information on the web page `http://www.aiml.net`.

logic called *Proof Theory* that deals with the structure of formal proofs, transformations between equivalent proofs, and the like. Deep results in Proof Theory include "cut-elimination theorems" telling us that – for appropriate logics – theorems derived in axiomatic format can also be derived in a "sequent calculus" (cf. Chapter 4) with only introduction rules for logical operators. The latter format is very perspicuous for theoretical purposes. Another proof-theoretic theme concerns the surplus of proofs, as finite combinatorial objects establishing validity. A proof often has "algorithmic content", which allows us to extract more concrete information about valid formulas, and perhaps even extract programs whose execution provably meets given specifications.

5.7 The completeness theorem

Perhaps the most important result for the minimal modal logic is this:[39] what K derives is the whole truth (*completeness*), and nothing but the truth (*soundness*).

Theorem 8. For all modal formulas φ, $\vdash_K \varphi$ iff $\models \varphi$

Completeness theorems were first proved by Emil Post for propositional logic in the 1920s, and – much deeper – by Kurt Gödel in 1929 for first-order logic. Surprisingly, they connect the very general notion of validity ($\models$) with a very concrete one of provability ($\vdash$).

Soundness. This is usually easy to prove, by induction on the length of proofs. One first checks that all axioms of the stated forms are valid, and next, that all the derivation rules preserve validity. For the minimal logic K, this is easy to see by inspection.[40]

Completeness. This involves a more complicated argument. It is like the completeness proofs for first-order logic that you may have seen already – but with some simplifications due to the simple structure of the modal language. The emphasis in this book will not be on proving completeness theorems, even though these are a large "industrial" part of the field. But we will tell you about the main proof steps here.

The cover argument. We argue by contraposition. Suppose a modal formula φ is not derivable in K. We reformulate this assumption using the following notion, that is important in its own right.

Definition 5.7.1 (Consistency)**.** A set Σ of formulas is *consistent* if for no finite conjunction σ of formulas from Σ, the negation $\neg\sigma$ is provable in the logic K.

[39] The result goes back essentially to Stig Kanger and Saul Kripke in the 1950s.

[40] Soundness is not always trivial: logics for program correctness have tricky proof rules for structured data and recursion, whose soundness can be in doubt.

Consistent sets have simple useful properties that we do not prove here. One is that, if φ is not derivable, then the set $\{\neg\varphi\}$ is consistent. Now we show that in general, any consistent set of formulas Σ has a satisfying model, which provides a semantic counter-example for the non-derivable formula φ, whence the latter is not semantically valid.

Maximally consistent sets. Any consistent set of formulas is contained in a *maximally consistent set of formulas*, i.e., a consistent set which has no consistent proper extensions. This may be proved by general set-theoretic principles ("Zorn's Lemma"). More popular is an explicit "Lindenbaum construction" enumerating all (countably many) formulas of the modal language: $\psi_1, \psi_2, \ldots$ and then, starting from Σ, working stage by stage, adding the currently scheduled formula if it is still consistent with those already chosen. In the countable limit, the result of this is still consistent, since by our definition, an inconsistency can only involve finitely many formulas, and hence it would already have shown up at some finite stage. It is easy to see that the preceding construction yields a maximally consistent set.

Maximally consistent sets have pleasant decomposition properties, making them behave like complete records of possible worlds:

Fact. Let Σ be a maximally consistent set. Then the following equivalences hold, for all modal formulas:

$$\begin{array}{lll} \text{(i)} & \neg\varphi \in \Sigma & \text{iff} \quad \text{not } \varphi \in \Sigma \\ \text{(ii)} & \varphi \wedge \psi \in \Sigma & \text{iff} \quad \varphi \in \Sigma \text{ and } \psi \in \Sigma \end{array}$$

This may be shown by simple propositional reasoning (K contains all Boolean tautologies). It also follows easily that maximally consistent sets are closed under K-derivable formulas.

Now we unpack modalities. First we say that $\Sigma R \Delta$ if for every formula α in the maximally consistent set Δ, we have $\Diamond\alpha$ in Σ. Soon, this R will become the accessibility relation in a model that we are going to create. Now we have the following further decomposition:

Fact. (iii) $\Diamond\varphi \in \Sigma$ iff there is some Δ with $\Sigma R \Delta$ and $\varphi \in \Delta$

The proof is trivial from right to left. From left to right, it is the only place in the whole completeness proof where we use the typically modal principles of K.[41] Consider the set of formulas $\Gamma = \{\varphi\} \cup \{\alpha \mid \Box\alpha \in \Sigma\}$: by the earlier definition of accessibility, any maximally consistent set containing this will be a R-successor of Σ.

Claim. The set Γ is consistent.

[41] In fact, here is where you could have *discovered* these key axioms!

Proof. Suppose it were not: then by the definition of consistency, there is a conjunction $\boldsymbol{\alpha}$ of some finite set of formulas with $\Box\alpha$ in Σ such that K proves $\neg(\boldsymbol{\alpha} \wedge \varphi)$, hence also $\boldsymbol{\alpha} \to \neg\varphi$. By Necessitation, K then also proves $\Box(\boldsymbol{\alpha} \to \neg\varphi)$, and by modal distribution, it proves $\Box\boldsymbol{\alpha} \to \Box\neg\varphi$. Now we assumed $\Box\alpha \in \Sigma$ for all separate $\alpha \in \boldsymbol{\alpha}$. But by our earlier observations about derivability, together, these formulas imply that $\Box\boldsymbol{\alpha} \in \Sigma$, and since maximally consistent sets are closed under provable consequence, we have $\Box\neg\varphi \in \Sigma$. Now, given K's definition of $\Diamond$ in terms of $\Box$, this contradicts the fact that $\Diamond\varphi \in \Sigma$. □

The Henkin model. Now we define a model $\boldsymbol{M} = (W, R, V)$ as follows.

Definition 5.7.2 (Canonical model)**.** The worlds W are all maximally consistent sets, the accessibility relation is the above defined relation R, and for the propositional valuation V, we set $\Sigma \in V(p)$ iff $p \in \Sigma$.

Then we have everything in place for the final argument:

Lemma (Truth Lemma)**.** For each maximally consistent set Σ, and each modal formula φ,

$$\boldsymbol{M}, \Sigma \models \varphi \quad \text{iff} \quad \varphi \in \Sigma$$

Proof. The proof is a straightforward induction on formulas φ, using all the ready-made ingredients provided in the decomposition facts for maximally consistent sets. A typical feature of the Truth Lemma, and one of its conceptual delights for logicians, is the harmony between a formula just *belonging* to Σ as a syntactic object, and that same formula being *true* at Σ, now viewed as one world in a universe of worlds where modal evaluation can take place. □

Actually, this proof establishes something more than we stated. "*Weak completeness*" is the property that every valid formula is derivable – or equivalently, that every consistent formula has a model. But we have really proved "*strong completeness*": all consistent sets have a model. Equivalently, this says that each valid consequence φ from a set of formulas Σ, finite or infinite, has a proof using only assumptions from that set: $\Sigma \models \varphi$ iff $\Sigma \vdash \varphi$.

There is something magical about this completeness argument, since we conjured up a counter-model out of our hat, using just the syntax of the modal language plus some simple combinatorial facts about provability in the minimal modal logic. Even so, the Henkin model *is* a concrete semantic object when we disregard the syntactic origins of its worlds, and as such it even has a remarkable property. All consistent sets can be made true in one and the same model! While we could also have seen this differently,[42] it is a characteristic fact about the modal language, that makes the Henkin model "largest" or "universal" among all models. By contrast, first-order logic has no such universal model, since its maximally consistent theories hate each other so much that they cannot live consistently in the same model.

Modern finite versions The following observation has become standard in the literature. One further remarkable feature of the completeness proof is that it can also be carried out in a *finite universe of formulas*. Just consider the initially given consistent formula and all its sub-formulas as your total language. Everything we have defined and proved also applies when relativized to that setting, with some notions and results even getting simpler. As a result, we get finite models for consistent formulas whose size is a function of the number of sub-formulas. This is one more way of seeing that modal logic is decidable.

5.8 Applications of completeness

You need to understand how completeness theorems are used. Note that $\models$ is defined with a universal quantifier over models, and $\vdash$ with an existential quantifier over proofs. This highlights one basic feature. If we want to show that some formula is non-derivable, this is a hard task proof-theoretically, as we need to see that *every* proof fails. But on the equivalent semantic side, this says there *exists* some model for the negation of the formula. So, one counter-example suffices. Note that this application involves only *soundness*.[43] Here is a more theoretical application of *completeness*. We show that some derivation rules beyond those stated are "admissible" for modal reasoning:

Fact. If $\vdash_K \Box\psi$, then $\vdash_K \psi$.

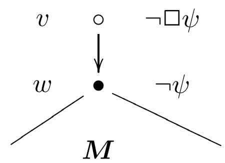
In other words, provability in K also satisfies a converse rule of "De-Necessitation". This rule is trivial in modal logics upward from T: but for K, it is by no means obvious.

Proof. Suppose that ψ is not provable. Then by completeness, there is a counter-model $\boldsymbol{M}$ with a world w where $\neg\psi$ holds. Now here is a semantic trick that is used a lot in modal logic. Take any *new* world v,

[42] For instance, here is an alternative construction. For each satisfiable set of formulas in the language, take one verifying model, and then form the so-called "*disjoint union*" of all these models: this also works.

[43] It is a tragic feature of many deep logical results that their easy side makes for the most concrete applications, whereas the deeper converse side is only there for theoretical enlightenment. Or stated more positively, the most important insights in logic are free from any base motive of practical gain.

add it to $\boldsymbol{M}$ and put just one extra R-link, from v to w:



The atomic valuation at v does not matter. In the new model $\boldsymbol{M}^+$, $\Box\psi$ is clearly false at v – hence it is not universally valid, and so by soundness, $\Box\psi$ is not derivable in K. $\Box$

In addition to axioms, modal rules of inference are a fascinating subject of study by themselves. A nice general result is "Rybakov's Theorem" from the 1980s, stating that it is *decidable* if a given rule of inference is admissible for the widely used modal logic *S4*.

Finally, having a complete axiom system does not make a logic decidable, witness the case of first-order logic. Given any formula φ, enumerating all possible proofs for it must indeed produce a proof if it is valid. But if φ is not valid, we have to sit through the entire infinite process, to make sure that it was not derivable. As it happens, we have seen already that modal validity was decidable – but this feature of modal logic was for additional reasons.

5.9 Coda: modal logic via proof intuitions

Purely proof-theoretic modal intuitions may run deep. See the passage on H.B. Smith in the *Manual of Intensional Logic* (van Benthem, 1988a). This early pioneer of modal logic in the 1930s took the view that the heart of modal reasoning was not in specific axioms or rules, but rather in two major principles which had to be respected by any modal system. Smith considered finite sequences of modalities as the core notions that a modal logic is trying to capture:

$$-, \Diamond, \Box, \Diamond\Diamond, \Diamond\Box, \Box\Diamond, \Box\Box, \ldots$$

He then formulated two intuitive desiderata on any concrete modal proof system:

(a) *Distinction*: no two distinct sequences are provably equivalent,
(b) *Comparison*: of any two sequences, one implies the other.

The two requirements are at odds, since (a) wants the logic to be weak, while (b) wants it to prove a lot. Indeed, these requirements cannot be met in "normal modal logics" extending K.[44] But even so,

[44] The reader may want to try her hand at the simple but clever argument.

Smith's intuitions are fascinating, and they have been modeled in suitably generalized logics over variants of our possible worlds semantics.

This excursion is also interesting because it reminds you how, through no fault of theirs, interesting ideas that do not "fit" may drop by the wayside as a science progresses.

Exercises Chapter 5

1. Prove the following formula in K: $\Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$.
2. Prove the formula $\Box\Diamond\Box\Diamond\varphi \leftrightarrow \Box\Diamond\varphi$ in the logic *K4*: *S4* minus Veridicality.
3. Show that the following rule is admissible in K: if $\vdash \Box\alpha \vee \Box\beta$, then $\vdash \alpha$ or $\vdash \beta$.
4. Prove the Boolean decomposition facts about maximally consistent sets stated in our completeness argument.
5. Prove that no "normal modal logic", i.e., a set of modal formulas extending the set of theorems of K and closed under Modus Ponens, Necessitation, and Substitution of formulas for proposition letters, satisfies both requirements stated by H. B. Smith.