

4 Models and Proofs

4.1 Soundness and completeness

In the last two chapters, we have met several notions of validity: K-validity, S5-validity, T-validity, etc. All these notions were defined in terms of possible-worlds models. A sentence is K-valid iff it is true at all worlds in all Kripke models; a sentence is T-valid iff it is true at all worlds in all reflexive Kripke models; and so on.

If you want to show that a sentence is, say, K-valid, you could in principle work through the clauses of definition 3.2 to show that the sentence is true at every world in every Kripke model. The tree method provides an alternative, and usually easier, means to check whether a sentence is valid. If all branches of a tree close, the target sentence is valid; if not, it is invalid.

Or so I claimed. But these claims aren't obvious. In this section, we are going to prove them. That is, we will prove the following two facts.

1. If all branches on a K-tree close, then the target sentence is K-valid.
2. If some branch on a fully expanded K-tree for a sentence remains open, then the target sentence is not K-valid.

By a *K-tree* I mean a tree that conforms to the K-rules from the previous chapter. By a *fully expanded* tree I mean a tree on which no more rules can be applied. (A fully expanded tree may be infinite.)

The first claim establishes the **soundness** of the tree method for K. In general, a proof method is called *sound* (with respect to a particular concept of validity) if everything that is provable with the method is valid.

The second claim establishes the **completeness** of the tree method for K. A proof method is called *complete* (with respect to a concept of validity) if everything that is valid is provable with the method. Equivalently: anything that is not provable is not valid.

Let's begin with soundness. I first give a rough outline of the proof idea, then I'll fill in the details.

We want to show that if a K-tree closes, then the target sentence is K-valid. So assume we have created a tree for a target sentence A and found that all branches close. We need to show that A is K-valid. To this end, we suppose for reductio that A is *not* K-valid. We will show that this leads to a contradiction.

By definition 3.3, a sentence is K-valid iff it is true at all worlds in all Kripke models. Our supposition that A is not K-valid therefore means that A is false at some world in some Kripke model. Let's call that world ' w ' and the model ' M '. Note that the closed tree begins with

$$1. \quad \neg A \quad (w)$$

So if we take the world variable ' w ' on the tree to pick out world w in M , then node 1 is a correct statement about M , insofar as $\neg A$ is indeed true at w in M . Now we can show the following:

If all nodes on some branch of a tree are correct statements about M , and the branch is extended by the K-rules, then all nodes on at least one of the resulting branches are still correct statements about M .

Since our complete tree is constructed from node 1 by applying the K-rules, it follows that on some branch of the tree, all nodes are correct statements about M . But the tree is closed. So every branch of the tree contains a pair of contradictory statements. These statements can't be correct statements about M . That's our contradiction.

Now let's fill in the details. First, let's say that a node on a tree is *a correct statement about* a model $M = (W, R, V)$ iff there is a function f that maps world variables to worlds in W such that either

- the node has the form $\omega R \nu$ and the R holds between $f(\omega)$ and $f(\nu)$, or
- the node has the form $A(\omega)$ and A is true at $f(\omega)$ in M .

We now prove the italicised statement above:

Soundness Lemma

If all nodes on some branch b of a tree are correct statements about a model M , and the branch is extended by applying a K-rule, then all nodes on at least one of the resulting branches are still correct statements about M .

Proof: We have to go through all the K-rules. In each case we assume that the rule is applied to some node(s) on some branch b , all nodes on which are correct statements about M . We show that on at least one of the resulting branches, all nodes are still correct statements about M .

- Suppose b contains a node of the form $A \wedge B (\omega)$ and the branch is extended by two new nodes $A (\omega)$ and $B (\omega)$. By assumption, $A \wedge B (\omega)$ is a correct statement about M . That is, there is some function f from world variables to worlds in M such that $M, f(\omega) \models A \wedge B$. By clause (c) of definition 3.2, it follows that $M, f(\omega) \models A$ and $M, f(\omega) \models B$. So the newly added nodes are also correct statements about M .
- Suppose b contains a node of the form $A \vee B (\omega)$ and the branch is split into two, with $A (\omega)$ appended to one end and $B (\omega)$ to the other. By assumption, there is some function f such that $M, f(\omega) \models A \vee B$. By clause (d) of definition 3.2, it follows that either $M, f(\omega) \models A$ or $M, f(\omega) \models B$. So at least one of the newly added nodes is also a correct statement about M .
- The proof for the other non-modal rules is similar. Let's move on to the rules for modal operators.
- Suppose b contains nodes of the form $\Box A (\omega)$ and $\omega R \nu$, and the branch is extended by adding $A (\nu)$. By assumption, there is some function f such that $M, f(\omega) \models \Box A$ and $f(\omega) R f(\nu)$. By clause (g) of definition 3.2, it follows that $M, f(\nu) \models A$. So the newly added node is also a correct statement about M .
- Suppose b contains a node of the form $\Diamond A (\omega)$ and the branch is extended by adding nodes $\omega R \nu$ and $A (\nu)$, where ν is new on the branch. By assumption, there is some function f such that $M, f(\omega) \models \Diamond A$. By clause (h) of definition 3.2, it follows that $M, \nu \models A$ for some ν in M such that $f(\omega) R \nu$. Let f' be the same as f except that $f'(\nu) = \nu$. Then the newly added nodes are correct

statements about M , using f' to interpret the world variables. Moreover, since f and f' differ at most for v , and v didn't occur on b before the addition of ωRv and $A(v)$, all earlier nodes on b are also correct statements about M if we use f' to interpret the world variables.

- The cases for $\neg\Box$ and $\neg\Diamond$ are similar to the previous two cases. □

With the help of the lemma we have just established, we can now prove that the method of K-trees is sound.

Theorem: Soundness of the tree rules for K

If a K-tree for a target sentence closes, then that sentence is K-valid.

Proof: We assume for reductio that the tree for some target sentence A closes even though A is not K-valid. If A is not K-valid, then $\neg A$ is true at some world w in some Kripke model M . So the first node on the tree, $\neg A(w)$, is a correct statement about M . Since the completed tree is created from this starting point by applying the K-rules, the Soundness Lemma implies that some branch on the tree only contains correct statements about M . But the tree closes, meaning that all its branches contain contradictory nodes of the form

- n. $B(v)$
- m. $\neg B(v)$

These two nodes can't both be correct statements about M . So we've reached a contradiction from the assumption that the tree for A closes even though A is not K-valid. □

Exercise 4.1

Fill in the cases for $A \rightarrow B$ and $\neg\Diamond$ in the proof of the Soundness Lemma.

Let's turn to completeness. We will show that if some branch on a fully expanded K-tree for a sentence remains open, then the tested sentence is not K-valid: it is false at some world in some Kripke model. We already know that a relevant countermodel can be read off from the open branch. We only need to prove that this method for generating countermodels always works.

First we define the method precisely.

Definition 4.1

For any open branch on any fully expanded tree, *the model induced by the branch* is the Kripke model (W, R, V) where:

- (a) W is the set of world variables on the branch,
- (b) $\omega R \nu$ holds in the model iff a node $\omega R \nu$ occurs on the branch,
- (c) for any sentence letter ρ and world ω , $V(\rho, \omega) = 1$ if $\rho(\omega)$ occurs on the branch, otherwise $V(\rho, \omega) = 0$.

We now show that all nodes on any open branch on a fully expanded tree are correct statements about the Kripke model induced by the branch.

Completeness Lemma

If b is an open branch on a fully expanded K-tree, and $M = \langle W, R, V \rangle$ is the model induced by b , then $M, \omega \models A$ for all sentences A and world variables ω for which $A(\omega)$ is on b .

The proof is by induction on the length of A . We first show that the lemma holds for sentence letters and negated sentence letters. Then we show that *if* the lemma holds for all sentences shorter than A (this is our *induction hypothesis*), *then* it also holds for A itself. It follows that the lemma holds for sentences of arbitrary length.

- If A is a sentence letter or a negated sentence letter, then the claim is true by clause (c) of definition 4.1 and clauses (a) and (b) of definition 3.2.
- If A is a doubly negated sentence $\neg\neg B$, then the branch also contains a node $B(\omega)$, because the tree is fully expanded. By induction hypothesis, $M, \omega \models B$. By clause (b) of definition 3.2, it follows that $M, \omega \models A$.
- If A is a conjunction $B \wedge C$, then the branch contains nodes for $B(\omega)$ and $C(\omega)$, because the tree is fully expanded. By induction hypothesis, $M, \omega \models B$ and $M, \omega \models C$. By clause (c) of definition 3.2, it follows that $M, \omega \models A$.

- If A is a negated conjunction $\neg(B \wedge C)$, then the branch contains either a node $\neg B$ (ω) or a node $\neg C$ (ω), because the tree is fully expanded. By induction hypothesis, $M, \omega \models \neg B$ or $M, \omega \models \neg C$. Either way, clauses (b) and (c) of definition 3.2 imply that $M, \omega \models A$.
- For the case where A is a disjunction, a conditional, a biconditional, or a negated disjunction, conditional, or biconditional, the proof is similar to one (or both) of the previous two cases.
- If A has the form $\Box B$, then the branch contains a B (ν) node for each world variable ν for which $\omega R\nu$ is on the branch (because the tree is fully expanded). By induction hypothesis, $M, \nu \models B$, for each such ν . By definition 4.1, it follows that $M, \nu \models B$ for all worlds ν such that $\omega R\nu$. By clause (g) of definition 3.2, it follows that $M, \omega \models \Box B$.
- If A has the form $\Diamond B$, then there is a world variable ν for which $\omega R\nu$ and B (ν) are on the branch (because the tree is fully expanded). By induction hypothesis, $M, \nu \models B$. And by definition 4.1, $\omega R\nu$. By clause (h) of definition 3.2, it follows that $M, \omega \models \Diamond B$. □
- For the case where A has the form $\neg\Box B$ or $\neg\Diamond B$, the proof is similar to one of the previous two cases.

Let's spell out how this establishes the completeness of the tree method.

Theorem: Completeness of the tree rules for K

If a sentence is K-valid, then there is a closed K-tree for that sentence.

Proof: Assume some sentence A is K-valid. Suppose for reductio that there is no closed K-tree for A . Take any open branch on any fully expanded tree for A . By the Completeness Lemma, A is false at w in the model induced by that branch. So A is not true at all worlds in all Kripke models. Contradiction. □

Exercise 4.2

Fill in the cases for $A \rightarrow B$ and $\neg\Diamond A$ in the proof of the Completeness Lemma.

So the tree rules for K are adequate: they allow us to prove all and only the K-valid sentences. The tree rules for S5 are also adequate, as are the rules for the various other systems we have looked at. In each case, the soundness and completeness proofs are mostly analogous to the proofs we just gave for K.

To illustrate what may have to be changed for other systems, let's have a brief look at the tree rules for T. Here we have an additional rule: the Reflexivity rule. To show that the T-rules are sound for the concept of T-validity, we can use the same strategy as above; we only need to check that the Soundness Lemma goes through for the case of the Reflexivity rule. That is, we need to check that if a branch contains only correct statements about a reflexive model M , and the branch is extended by the Reflexivity rule, then the added node is still a correct statement about M . This is evidently the case.

For completeness, we can also use the same strategy as above. The Completeness Lemma will tell us that if a T-tree remains open, then the target sentence is false at world w in the Kripke model induced by any open branch on the tree. We only need to check that this induced model is reflexive. And it must be, because an open branch on a fully expanded T-tree contains $\omega R \omega$ for each world variable ω on the branch.

Exercise 4.3

When constructing a tree proof, one often has a choice of which rules to apply in which order. Show that the choice doesn't matter, in the following sense: If there is a K-tree for the target sentence in which all branches close, then there is no fully expanded K-tree for the target sentence in which some branch remains open.

4.2 Axiomatic proofs

The tree method is easy to use. The older method of axiomatic proofs is not. The method is nonetheless worth studying, and not only because of its historical importance.

An axiomatic ("Hilbert-style") proof is a sequence of sentences each of which is either an axiom or follows from earlier sentences in the sequence by a rule.

In order to reduce the required axioms and rules, it is customary to use a limited version of \mathcal{L}_M in which the only logical operators are \neg , \rightarrow , and \Box . This is no real

loss, since any sentence with other operators can be transformed into an equivalent sentence with only \neg , \rightarrow , and \Box , by the following equivalences:

$$\begin{aligned} A \wedge B &\Leftrightarrow \neg(A \rightarrow \neg B) \\ A \vee B &\Leftrightarrow \neg A \rightarrow B \\ A \leftrightarrow B &\Leftrightarrow \neg((A \rightarrow B) \rightarrow \neg(B \rightarrow A)) \\ \Diamond A &\Leftrightarrow \neg\Box\neg A \end{aligned}$$

A well-known axiomatic calculus for classical propositional logic consists of the following three axiom schemas, **A1**–**A3**, together with the rule of *Modus Ponens*, **MP**.

- (A1) $A \rightarrow (B \rightarrow A)$
- (A2) $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$
- (A3) $(\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$
- (MP) If A and $A \rightarrow B$ occur on a proof, you may append B .

Each axiom schema has infinitely many instances, all of which count as axioms.

Any sentence that is valid in classical propositional logic can be derived from some instances of A1–A3 with the use of MP. To illustrate, here is a proof of $p \rightarrow p$, with added annotations.

1. $p \rightarrow ((p \rightarrow p) \rightarrow p)$ (A1)
2. $(p \rightarrow ((p \rightarrow p) \rightarrow p)) \rightarrow ((p \rightarrow (p \rightarrow p)) \rightarrow (p \rightarrow p))$ (A2)
3. $(p \rightarrow (p \rightarrow p)) \rightarrow (p \rightarrow p)$ (1, 2, MP)
4. $p \rightarrow (p \rightarrow p)$ (A1)
5. $p \rightarrow p$ (3, 4, MP)

To get a complete axiomatization for system **K**, we only need one more axiom schema and one more rule. The axiom schema is **K**:

$$\text{(K)} \quad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

The new rule is called “necessitation”:

(Nec) If A occurs on a proof, you may append $\Box A$.

In general, of course, one can’t infer $\Box A$ from A . After all, the sentence A may be true (at a world) while $\Box A$ is false (at the same world). However, the **Nec** rule is not a rule for drawing inferences from arbitrary assumptions. It allows you to infer $\Box A$ only if A already occurs earlier in the same proof. And everything that occurs in an axiomatic proof is either an axiom or it follows from the axioms by the rules. Given that the axioms are logical truths, and given that whatever logically follows from a logical truth itself a logical truth, any sentence that occurs on some line in an axiomatic proof is a logical truth. In effect, **Nec** therefore assumes that if a sentence A is logically true, then $\Box A$ is also logically true.

To illustrate the use of **K** and **Nec**, let’s look at a proof of $\Box(p \wedge q) \rightarrow \Box p$. In our restricted version of \mathcal{L}_M , this is $\Box\neg(p \rightarrow \neg q) \rightarrow \Box p$. The proof begins with a proof of the tautology $\neg(p \rightarrow \neg q) \rightarrow p$, using **A1–A3** and **MP**. This takes about 60 lines, which we don’t need to worry about. From then, we proceed as follows.

- | | | |
|-----|--|--------------------|
| 60. | $\neg(p \rightarrow \neg q) \rightarrow p$ | (from A1–A3 by MP) |
| 61. | $\Box(\neg(p \rightarrow \neg q) \rightarrow p)$ | (60, Nec) |
| 62. | $\Box(\neg(p \rightarrow \neg q) \rightarrow p) \rightarrow (\Box\neg(p \rightarrow \neg q) \rightarrow \Box p)$ | (K) |
| 63. | $\Box\neg(p \rightarrow \neg q) \rightarrow \Box p$ | (61, 62, MP) |

All and only the **K**-valid sentences can be derived from **A1–A3** and **K** by **MP** and **Nec**. Adding further axioms (i.e., axiom schemas) yields axiomatic calculi for stronger concepts of validity. For example, if we add the **T** schema $\Box A \rightarrow A$, we get an axiomatic calculus for system **T**: all and only the **T**-valid sentences can be derived from **A1–A3**, **K**, and **T** by **MP** and **Nec**. Adding the **4** schema $\Box A \rightarrow \Box\Box A$ to the calculus for **T** gives us a calculus for **S4**. Alternatively, adding the **5** schema $\Diamond A \rightarrow \Box\Diamond A$ to the calculus for **T** gives us a calculus for **S5**.

Exercise 4.4

Outline a proof of $\Box p \rightarrow \Diamond p$ in the axiomatic calculus for **T**. You can assume that any propositional tautology is somehow derivable from **A1–A3** and **MP**,

without giving its proof.

Exercise 4.5

Suppose we add the (“McKinsey”) schema $\Box\Diamond A \rightarrow \Diamond\Box A$ to the axiomatic calculus for S5. Explain why we can then prove all instances of the (“Triv”) schema $\Box A \leftrightarrow A$. As in the previous exercise, you can assume that any propositional tautology is provable.

When I say that all and only the K-valid sentences are provable in the axiomatic calculus for K, I am making two claims:

1. Everything that’s provable in the axiomatic calculus for K is K-valid. (Soundness)
2. Everything that’s K-valid is provable in the axiomatic calculus for K. (Completeness)

Soundness is relatively easy to prove. We have to show that anything that is derivable from (some instances of) **A1–A3** and **K** by **MP** and **Nec** is true at all worlds in all Kripke models. To this end, we first show that all the axioms are true at all worlds in all Kripke models. Then we show that if a sentence is derived from other sentences by **MP** or **Nec**, and the other sentences are true at all worlds in all Kripke models, then the derived sentence is also true at all worlds in all Kripke models. In other words, we show that the axioms are valid, and that the rules preserve validity. It follows that anything that can be derived from the axioms by the rules is valid.

Theorem: Soundness of the axiomatic calculus for K

If A is provable in the axiomatic calculus for K, then A is K-valid.

Proof: We first show that all the axioms are K-valid.

A1 By clause (e) of definition 3.2, an instance of $A \rightarrow (B \rightarrow A)$ is *false* at some world w in some Kripke model M iff (the corresponding instances of) A and B are both true at w while A is false at w . But we can’t have both $M, w \models A$ and $M, w \not\models A$. So the assumption that some instance of $A \rightarrow (B \rightarrow A)$ is false at some world in some Kripke model has led to a contradiction. So every instance

of the schema is true at every world in every Kripke model.

A2 By clause (e) of definition 3.2, an instance of $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$ is false at some world w in some Kripke model M iff (the corresponding instances of) $A \rightarrow (B \rightarrow C)$, $A \rightarrow B$, and A are true at w while C is false at w . However, clause (e) also implies that if $A \rightarrow B$ and A are both true at w , then so is B . Similarly, clause (e) implies that if $A \rightarrow (B \rightarrow C)$ and A and B are true at w , then so is C . We have reached a contradiction: $M, w \models C$ and $M, w \not\models C$. So every instance of $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$ is true at every world in every Kripke model.

A3 Exercise.

K See observation 3.1 on p.53.

Now we show that **MP** and **Nec** preserve validity.

MP By clause (e) of definition 3.2, if some instance of $A \rightarrow B$ is true at some world in some Kripke model, and so is (the corresponding instance of) A , then B is also true at that world in that model. So if $A \rightarrow B$ and A are true at *all* worlds in *all* Kripke models, then B is also true at all worlds in all Kripke models.

Nec By clause (g) of definition 3.2, $\Box A$ is true at a world w in a Kripke model iff A is true at all worlds accessible from w . It follows that if A is true at *all* worlds in all Kripke models, then $\Box A$ is also true at all worlds in all Kripke models.

□

Exercise 4.6

Fill in the missing case for **A3**.

The above proof is easily extended to other modal systems, defined in terms of some restricted class of Kripke models. As we've just shown, **A1–A3** and **K** are true at all worlds in all Kripke models. So they are also true at all worlds in any more restricted class of Kripke models. Moreover, the above arguments for **MP** and **Nec** really show that these rules preserve validity in any class of Kripke models. So if we want to show that, say, the axiomatic calculus for T is sound with respect to the

concept of T-validity – that is, if we want to show that anything that is derivable from **A1–A3**, **K**, and **T** by **MP** and **Nec** is true at all worlds in all reflexive Kripke models – all that is left to do is to show that every instance of the **T**-schema is true at all worlds in all reflexive Kripke model. (We’ve already shown this: see observation 3.3.)

Exercise 4.7

Outline the soundness proof for the axiomatic calculus for S4.

4.3 Canonical models

Next, completeness. We are going to show that any K-valid sentence is derivable from some instances of **A1–A3** and **K** by the rules **MP** and **Nec**. The argument is rather complicated, but it reveals some independently interesting facts about systems of modal logic.

Like in the case of the tree method, we will argue by contraposition. We will show that any sentence that *cannot* be derived from **A1–A3** and **K** by **MP** and **Nec** is *not* K-valid. To show that a sentence is not K-valid, we will give a countermodel: a Kripke model in which the sentence is false at some world. In fact, we will give the *same* countermodel for every sentence that isn’t derivable in the calculus. You might think we need different countermodels for different sentences, but it turns out that we don’t. There is a particular model in which every K-invalid sentence is false at some world. This model is called the **canonical model** for system K, or for the corresponding axiomatic calculus.

Before I define the canonical model, let me introduce some shorthand terminology. (We’ll only use these terms in the present section.) We’ll say that an \mathcal{L}_M sentence is *K-provable* if it can be proved in the axiomatic calculus for K. A set of \mathcal{L}_M -sentences will be called *K-inconsistent* if it contains a finite number of sentences A_1, \dots, A_n such that $\neg(A_1 \wedge \dots \wedge A_n)$ is K-provable. A set is *K-consistent* if it is not K-inconsistent.

(For example, the set $\{\Box(p \wedge q), q \rightarrow p, \neg\Box q\}$ is K-inconsistent, because it contains two sentences, $\Box(p \wedge q)$ and $\neg\Box q$ whose conjunction is refutable in K, in the sense that the negation $\neg(\Box(p \wedge q) \wedge \neg\Box q)$ of their conjunction is derivable from some instances of **A1–A3** and **K** by **MP** and **Nec**.)

A set of \mathcal{L}_M -sentences is called *maximal* if for every \mathcal{L}_M -sentence, it contains either that sentence or its negation. A set is *maximal K-consistent* if it is both maximal and

K-consistent.

Now here's the canonical model for K.

Definition 4.2: Canonical model

The **canonical model** M_K for K is the Kripke model $\langle W, R, V \rangle$, where

- W is the set of all maximal K-consistent sets of \mathcal{L}_M -sentences.
- wRv iff v contains every sentence A for which w contains $\Box A$.
- For every sentence letter ρ and world w in W , $V(\rho, w) = 1$ iff $\rho \in w$.

So the “worlds” in the canonical model are sets of \mathcal{L}_M -sentences. The interpretation function makes a sentence letter true at a world iff the letter is an element of the world. As we are going to see, this generalizes to arbitrary sentences:

- (1) A world w in M_K contains all and only the sentences that are true at w in M_K .

We will also prove the following:

- (2) If some sentence cannot be proved in the axiomatic calculus for K, then its negation is a member of some world in M_K .

Together, these two lemmas will establish the completeness of the calculus. Recall that completeness means that if a sentence isn't K-provable in then there is some world in some Kripke model at which it is false. Fact (2) tells us that if a sentence A isn't K-provable, then $\neg A$ is a member of some world w in the canonical model M_K . By fact (1), we can infer that $\neg A$ is true at w in M_K . So A is false at w in M_K . So A is not K-valid.

We are going to prove (2) first. We'll need the following observation. (Feel free to skip the proof.)

Observation 4.1: If a set Γ is K-consistent, then for any sentence A , either $\Gamma \cup \{A\}$ or $\Gamma \cup \{\neg A\}$ is K-consistent. ($\Gamma \cup \{A\}$ is the set that contains all members of Γ as well as A .)

Proof sketch: Suppose for reductio that both $\Gamma \cup \{A\}$ and $\Gamma \cup \{\neg A\}$ are K-inconsistent. Since $\Gamma \cup \{A\}$ is K-inconsistent, there are sentences A_1, \dots, A_n in $\Gamma \cup \{A\}$ such

that $\neg(A_1 \wedge \dots \wedge A_n)$ is K-provable. Since Γ itself is K-consistent, one of the sentences A_1, \dots, A_n must be A . Let B be the conjunction of the other sentences in A_1, \dots, A_n , all of which are in Γ . So $\neg(B \wedge A)$ is K-provable.

Similarly, since $\Gamma \cup \{\neg A\}$ is K-inconsistent, there are sentences A_1, \dots, A_n in $\Gamma \cup \{\neg A\}$ such that $\neg(A_1 \wedge \dots \wedge A_n)$ is K-provable, and one of these sentences must be $\neg A$. Let C be the conjunction of the others, all of which are in Γ . So $\neg(C \wedge \neg A)$ is K-provable.

So both $\neg(B \wedge A)$ and $\neg(C \wedge \neg A)$ are K-provable. In propositional logic, $\neg(B \wedge A)$ and $\neg(C \wedge \neg A)$ entail $\neg(B \wedge C)$. And indeed, $\neg(B \wedge C)$ is derivable from $\neg(B \wedge A)$ and $\neg(C \wedge \neg A)$ through some applications of **A1–A3** and **MP**. So $\neg(B \wedge C)$ is K-provable. But $B \wedge C$ is a conjunction of sentences from Γ . So Γ is K-inconsistent. This contradicts the assumption that Γ is K-consistent. \square

Now we can prove fact (2), which is known as “Lindenbaum’s Lemma”.

Lindenbaum’s Lemma

Every K-consistent set is included in some maximal K-consistent set.

Proof: Let S_0 be some K-consistent set of sentences. Let A_1, A_2, \dots be a list of all \mathcal{L}_M -sentences in some arbitrary order. For every number $i \geq 0$, define

$$S_{i+1} = \begin{cases} S_i \cup \{A_i\} & \text{if } S_i \cup \{A_i\} \text{ is K-consistent} \\ S_i \cup \{\neg A_i\} & \text{otherwise.} \end{cases}$$

This gives us an infinite list of sets S_0, S_1, S_2, \dots . Note that each set in the list is K-consistent: S_0 is K-consistent by assumption. And if some set S_i in the list is K-consistent, then either $S_i \cup \{A_i\}$ is K-consistent, in which case $S_{i+1} = S_i \cup \{A_i\}$ is K-consistent, or $S_i \cup \{A_i\}$ is not K-consistent, in which case S_{i+1} is $S_i \cup \{\neg A_i\}$, which is K-consistent by observation 4.1. So if any set in the list is consistent, then the next set in the list is also consistent. So S_0, S_1, S_2, \dots are all K-consistent.

Now let S be the set of sentences that occur in at least one of the sets S_1, S_2, S_3, \dots . (That is, let S be the union of S_1, S_2, S_3, \dots .) Evidently, S_0 is included in S . And S is maximal. Moreover, S is K-consistent. For if S were not K-consistent, then it would contain some sentences B_1, \dots, B_n such that $\neg(B_1 \wedge \dots \wedge B_n)$ is K-provable. All of these sentences would have to occur somewhere on the list A_1, A_2, \dots . Let

A_j be a sentence from A_1, A_2, \dots that occurs after all the B_1, \dots, B_n . If B_1, \dots, B_n are in S , they would have to be in S_j already, so S_j would be K-inconsistent. But we've seen that all of S_0, S_1, S_2, \dots are K-consistent. So S is a maximal K-consistent set that includes S_0 . \square

To prove fact (1), we need another observation.

Observation 4.2: If Γ is a maximal K-consistent set of sentences that does not contain $\Box A$, and Γ^- is the set of all sentences B for which $\Box B \in \Gamma$, then $\Gamma^- \cup \{\neg A\}$ is K-consistent.

Proof sketch: We show that if $\Gamma^- \cup \{\neg A\}$ is not K-consistent, then neither is Γ . If $\Gamma^- \cup \{\neg A\}$ is not K-consistent, then there are sentences B_1, \dots, B_n in Γ^- such that $\neg(B_1 \wedge \dots \wedge B_n \wedge \neg A)$ is K-provable. And then $(B_1 \wedge \dots \wedge B_n) \rightarrow A$ is K-provable, because it is a tautological consequence of $\neg(B_1 \wedge \dots \wedge B_n \wedge \neg A)$. By repeated application of **Nec**, **K**, and **MP**, we can derive $(\Box B_1 \wedge \dots \wedge \Box B_n) \rightarrow \Box A$. By propositional logic, we get $\neg(\Box B_1 \wedge \dots \wedge \Box B_n \wedge \neg \Box A)$. So $\{\Box B_1, \dots, \Box B_n, \neg \Box A\}$ is K-inconsistent. But $\Box B_1, \dots, \Box B_n$ are in Γ^- . And since $\Box A$ is not in Γ and Γ is maximal, $\neg \Box A$ is in Γ . So $\{\Box B_1, \dots, \Box B_n, \neg \Box A\}$ is a subset of Γ . So Γ is K-inconsistent. \square

Here, then, is fact (1):

Canonical Model Lemma

For any world w in M_K and any sentence A , A is in w iff $M_K, w \models A$.

Proof: The proof is by induction on complexity. We first show that the lemma holds for sentence letters. Then we show that *if* the lemma holds for some sentences A and B , *then* it also holds for the more complex sentences $\neg A$, $A \rightarrow B$, and $\Box A$.

1. (Sentence letters.) If A is a sentence letter, then by definition 4.2, $V(w, A) = 1$ iff $A \in w$, and so by clause (a) of definition 3.2, $M_K, w \models A$ iff $A \in w$.
2. (Case $\neg A$.) By clause (b) of definition 3.2, $M_K, w \models \neg A$ iff $M_K, w \not\models A$. If the lemma holds for A , then $M_K, w \not\models A$ iff $A \notin w$. And because w is maximal, we have $A \notin w$ iff $\neg A \in w$. So $M_K, w \models \neg A$ iff $\neg A \in w$.

3. (Case $A \rightarrow B$.) $M_K, w \models A \rightarrow B$ iff either $M_K, w \not\models A$ or $M_K, w \models B$, by clause (c) of definition 3.2. Given that the lemma holds for A and B , we have $M_K, w \not\models A$ iff $A \notin w$, and $M_K, w \models B$ iff $B \in w$. So $M_K, w \models A \rightarrow B$ iff either $A \notin w$ or $B \in w$. And since w is maximal consistent, $A \rightarrow B \in w$ iff either $A \notin w$ or $B \in w$. So $M_K, w \models A \rightarrow B$ iff $A \rightarrow B \in w$.
4. (Case $\Box A$.) From left to right, suppose $\Box A \in w$. By definition 4.2, it follows that $A \in v$ for all v with wRv . Assuming that the lemma holds for A , we have $M_K, v \models A$ for all v with wRv . So $M_K, w \models \Box A$, by clause (g) of definition 3.2. For the converse direction, suppose $\Box A \notin w$. Let Γ^- be the set of all sentences B such that $\Box B \in w$. By observation 4.2, $\Gamma^- \cup \{\neg A\}$ is K-consistent. By definition 4.2 and Lindenbaum's Lemma, it follows that there is some $v \in W$ such that wRv and $\neg A \in v$. So $A \notin v$. Assuming that the lemma holds for A , we have $M_K, v \not\models A$. So $M_K, w \not\models \Box A$, by clause (g) of definition 3.2. \square

The completeness of the axiomatic calculus for K follows immediately from the previous two lemmas, as foreshadowed above:

Theorem: Completeness of the axiomatic calculus for K

If A is K-valid, then A is provable in the axiomatic calculus for K.

Proof: We show that if A is not K-provable then it is not K-valid. Suppose A is not K-provable. Then $\{\neg A\}$ is K-consistent. It follows by Lindenbaum's Lemma that $\{\neg A\}$ is included in some maximal K-consistent set S . By definition 4.2, that set is a world in M_K . Since $\neg A$ is in S , it follows from the Canonical Model Lemma that $M_K, S \models \neg A$. So $M_K, S \not\models A$. So A is not true at all worlds in all Kripke models. \square

Done!

Let me briefly explain how the completeness proof would have to be adjusted for other modal systems.

The definition of M_K is easily generalised to arbitrary axiomatic calculi: simply replace 'K-consistent' by 'T-consistent' or 'S4-consistent', etc. Lindenbaum's Lemma and the Canonical Model Lemma can also be generalised in this fashion. (The proof goes through just the same if you replace 'K-consistent' by 'T-consistent' or 'S4-consistent', etc.)

Now suppose we want to prove that all T-valid sentences can be proved in the axiomatic calculus for T, which has the additional axiom (schema) $\Box A \rightarrow A$. The argument starts as before. We suppose that some sentence A is not provable in the axiomatic calculus for T. Our goal is to show that A is not valid in the class of reflexive Kripke models.

If A is not T-provable, then $\{\neg A\}$ is T-consistent. It follows by Lindenbaum's Lemma that $\{\neg A\}$ is included in some maximal T-consistent set S . By definition of canonical models, that set is a world in the canonical model M_T for T. Since $\neg A$ is in S , it follows from the Canonical Model Lemma that $M_T, S \models \neg A$. So $M_T, S \not\models A$. This shows that there is some world in some Kripke model where A is false. But what we need to show is that there is some world in some *reflexive* Kripke model where A is false.

We can close this gap by showing that M_T is reflexive. By definition 4.2, a world w in a canonical model is accessible from itself iff whenever $\Box A \in w$ then $A \in w$. Since the worlds in M_T are maximal T-consistent sets of sentences, and every such set contains every instance of the **T** schema $\Box A \rightarrow A$, there is no world in M_T that contains $\Box A$ but not A . So every world in M_T has access to itself.

The last step goes through for the most common systems of modal logic, but sometimes it fails. Sometimes, an axiomatic calculus is sound and complete with respect to some class of Kripke models, but the canonical model of the calculus is not a member of that class. Completeness must then be established by some other means. An example is the system GL from the next section.

Exercise 4.8

Outline the completeness proof for the axiomatic calculus for S4.

4.4 Provability logic

We've looked at some proofs about what is and isn't provable with a certain method. When we prove facts about proofs, we are doing *meta-logic*. The meta-logic proofs we gave in the previous sections were informal: We did not use a formal (natural deduction or tree or axiomatic) method to prove soundness and completeness theorems. But in principle that could be done. In particular, the proofs could all be formalized in an axiomatic calculus for predicate logic with a few additional axioms about

sets. (We needed some assumptions about sets e.g. in the proof of Lindenbaum's Lemma.) A well-known calculus of that kind is ZFC (named after Ernst Zermelo, Abraham Fraenkel, and the Axiom of Choice). ZFC is strong enough to prove not just soundness and completeness in modal logic, but practically everything that can be proved in any branch of maths.

An interesting feature of ZFC is that it can prove facts about provability not just in weaker axiomatic calculi; it can also prove facts about provability in ZFC itself. For example, one can prove *in ZFC* that whenever $A \rightarrow B$ and A are both provable in ZFC, then so is B .

This brings us back to modal logic. Suppose we read the box as 'it is mathematically provable that . . .', and we understand mathematical provability as provability in ZFC. Then every instance of the **K**-schema

$$\mathbf{(K)} \quad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

in the language of ZFC is true. More interestingly, as I just mentioned, every such instance is provable in ZFC. (In the language of ZFC, the provability predicate is of course not written as a box. But for every instance of **K** there is an equivalent ZFC-sentence, if the box is read as 'provable in ZFC', that is provable in ZFC.)

One can also show that if a sentence A is provable in ZFC, then it is provable in ZFC that A is provable in ZFC. Moreover, since ZFC is an extension of classical predicate logic, every propositional tautology is provable in ZFC, and **MP** is an inference rule of ZFC.

So the logic of mathematical provability validates all the axioms and rules of the basic modal logic **K**: We have **A1–A3**, **K**, **Nec**, and **MP**. What other principles do we have?

You might expect that the **T**-schema should be provable:

$$\mathbf{(T)} \quad \Box A \rightarrow A$$

Surely if something is mathematically provable, then it is true. However, it turns out that there are instances of **T** that aren't provable. That is, while every instance of **T** is intuitively true, we have no mathematical proof of this assumption. Instead, the following shocking principle can be established (in ZFC).

$$\mathbf{(GL)} \quad \Box(\Box A \rightarrow A) \rightarrow \Box A$$

The schema is named after Kurt Gödel and Martin Löb. The system **GL**, which is axiomatized by **A1–A3**, **K**, **GL**, **Nec**, and **MP**, completely captures what ZFC can prove about provability in ZFC.

To see why **GL** is shocking, consider an arbitrary mathematical falsehood – say, $2+2=5$. Obviously, $2+2=5$ is not provable in ZFC. I say “obviously”, but can we prove (in ZFC) that $2+2=5$ is not provable? The answer is no. Here’s why. If we could prove $\neg\Box(2+2=5)$, then we could also prove $\Box(2+2=5) \rightarrow (2+2=5)$, which is a truth-functional consequence of $\neg\Box(2+2=5)$. And since every instance of **GL** is provable, if we could prove $\Box(2+2=5) \rightarrow (2+2=5)$ then we could also prove $2+2=5$ itself. So if we could prove that we can’t prove that $2+2=5$, then we could also prove that $2+2=5$!

What we see here is a reflection of a deep fact about the limitations of mathematical provability, first revealed by Kurt Gödel. Gödel’s “second incompleteness theorem” states that no consistent axiomatic system that is powerful enough to formalize elementary mathematical reasoning (and for which there is a finite procedure for testing whether a sentence is an axiom) can prove its own consistency. Note that an inconsistent axiomatic system (in classical logic) can prove everything, because everything follows from a contradiction. So if ZFC could prove that it *can’t* prove $2+2=5$, then ZFC could in effect prove its own consistency. Gödel’s theorem tells us that ZFC would then be inconsistent – in which case it could prove anything, including $2+2=5$.

Surprisingly, Kripke models play an important role in the study of mathematical provability: the standard technique for proving that all instance of **GL** are provable in ZFC draws on the fact that **GL** is sound and complete with respect to the class of finite, transitive, and irreflexive Kripke models. This is surprising because intuitively, mathematical truths are true at all possible worlds, so it is hard to see how mathematical provability could be usefully analysed in terms of truth at accessible worlds.

Exercise 4.9

Explain why the logic of provability does not contain all instances of **5**, given that it contains all instances of **GL**.