ADAM BJORNDAHL

MODAL LOGIC COURSE NOTES

VERSION: APRIL 19, 2020

The Plan

THESE COURSE NOTES are designed to form the basis for the second half of the *Modal Logic* (80-315/615) course offered through the Philosophy Department at Carnegie Mellon University.

All in-person classes were cancelled over spring break due to the COVID-19 outbreak, and like many institutions, we have transitioned to fully online instruction for the remainder of the semester.

- These notes will be freely available to all students in the class, updated with new material on a weekly basis.
- At the **beginning of each week**, I will release a pre-recorded "lecture" explaining and supplementing the new additions to the course notes.
- Each **Tuesday from 15:00 to 16:20**,[1] there will be a live Zoom session open to all students to discuss the new material and address any other questions.
- Each **Thursday** a Canvas-based quiz ("quest") will be made available (all day) consisting of short-answer comprehension questions covering the most recent course content. Quests will be 30-minutes long.
- Problem Sets will continue to be assigned every two weeks according to the following schedule:[2]
  - March 31: Problem Set 4 due
  - April 2: Problem Set 5 assigned
  - April 14: Problem Set 5 due
  - April 16: Problem Set 6 assigned
  - April 28: Problem Set 6 due
  - Problem Set 7 is cancelled.

[1] Eastern daylight time.

[2] This is delayed by 1 week from what's was on the syllabus.

Modal Syntax and Semantics

Modal logics are expansions of classical logic designed for reasoning about a richer variety of phenomena. While classical propositional logic can connect statements[3] together with simple logical relations,[4] it cannot capture the more complex relationships between statements like

[3] E.g., "It's raining" or "John is wearing a blue hat" or "The value of x is 3".
[4] E.g., "and", "or", "not", "if...then" (the material conditional), and "if and only if".

- "It's raining" and "Alice knows that it's raining";
- "The value of x is 3" and "The value of x will continue to be 3";
- "John is wearing a blue hat" and "If John were wearing a red hat, he wouldn't know it".

The **basic propositional**[5] **modal language** is specified recursively as follows:[6]

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi,$$

where $p \in$ PROP, a countable set of *primitive propositions*. Other familiar logical connectives like $\vee$ and $\rightarrow$ are defined as abbreviations in the usual way.[7] We also write $\Diamond\varphi$ as an abbreviation for $\neg\Box\neg\varphi$.[8]

[5] I.e., without quantifiers.

[6] We begin by restricting our attention to the simplest language we can, so as to make most salient the particular phenomena we aim to investigate. We can always expand the language later, as needed.

[7] Specifically, $\varphi \vee \psi$ is an abbreviation for $\neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi$ is an abbreviation for $\neg(\varphi \wedge \neg\psi)$, and $\varphi \leftrightarrow \psi$ is an abbreviation for $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$.
[8] This is called the "dual" of $\Box$.

The symbol $\Box$ is chosen to look "neutral", corresponding to the intention for $\Box\varphi$ to be open to a variety of interpretations (depending on the application).[9] What unites these seemingly disparate interpretations is that they can all be analyzed, mathematically, as a type of *bounded universal quantification*.[10] For instance, "Alice knows φ" can be interpreted as saying something like "φ is true in all configurations of the world that Alice currently considers possible". For another example, "φ will be true henceforth" might naturally be interpreted as saying that "φ is true at every point of time in the future". Note that in both cases, the modality is analyzed using words like "all" or "every".

[9] E.g., "φ is known", "φ is believed", "φ will be true henceforth", "φ ought to be the case", "φ is a guaranteed result (of taking some action)", "φ is knowable", etc.
[10] That is, universal quantification over some restricted set.

These ideas are formalized in the definitions of *frames* and *models*, and the associated *semantics* for formulas in the basic modal language.

Frames and models

A **frame** is a pair (X, R) where X is nonempty set[11] and $R \subseteq X \times X$ is a binary *accessibility relation* on X. We write xRy to indicate that $(x, y) \in R$ and in this case say that *y is accessible from x*. We also define $R(x) = \{y \in X : xRy\}$, the set of all worlds accessible from x.

[11] Elements of which are called "worlds", or "states", or "points", etc.

A **model** (over PROP) is a frame $F = (X, R)$ together with a *valuation function* $v : \text{PROP} \to 2^X$.[12] The idea is that a valuation function specifies all the (contingent) truth values of each primitive proposition $p \in \text{PROP}$ at each world $x \in X$—when $x \in v(p)$ we say that *p is true at x*.

[12] The notation 2^X denotes the powerset of X, that is, the set of all subsets of X.

This world-dependent notion of truth in a model $M = (F, v)$ is extended to all formulas in the basic modal language via the following recursive semantic clauses:

$$\begin{aligned}
[\![p]\!]_M &= v(p) \\
[\![\neg\varphi]\!]_M &= X \setminus [\![\varphi]\!]_M \\
[\![\varphi \wedge \psi]\!]_M &= [\![\varphi]\!]_M \cap [\![\psi]\!]_M \\
[\![\Box\varphi]\!]_M &= \{x \in X : R(x) \subseteq [\![\varphi]\!]_M\}.
\end{aligned}$$

The set $[\![\varphi]\!]_M$ is sometimes called the *truth set* or the *extension* of φ in M; it is the set of worlds in M where φ is true. We also write $(M, x) \models \varphi$ for $x \in [\![\varphi]\!]_M$, and in this case say that *φ is true at x*.[13] In particular, $\Box\varphi$ is true at x if and only if φ is true at every world accessible from x.[14]

[13] When the model is clear from context we sometimes omit it from the notation, writing for example $[\![\varphi]\!]$ or $x \models \varphi$.

[14] It is also not hard to check that $\Diamond\varphi$ is true at x just in case there exists a world y accessible from x where φ is true.

One easily checks that the recursive clauses for $\neg$ and $\wedge$ guarantee that each world $x \in X$ behaves like a classical model as far as Boolean connectives are concerned.[15]

[15] That is, the standard truth table definitions of truth for the Boolean connectives hold at each world.

A formula φ is said to be **valid in a model** M if $[\![\varphi]\!]_M = X$; that is, if φ is true at *every* world in M. In this case we write $M \models \varphi$. A formula φ is said to be **valid in a frame** F if $M \models \varphi$ for every model M based on F.[16] In this case we write $F \models \varphi$. A formula φ is said to be **valid in a class of frames** $\mathcal{F}$ if $F \models \varphi$ for all $F \in \mathcal{F}$. In this case we write $\mathcal{F} \models \varphi$. And finally, a formula φ is said to be simply **valid** if it is valid on every frame. In this case we write $\models \varphi$.

[16] A model (X, R, v) is said to be *based on* the frame (X, R).

We have seen in class many examples of formulas that are or are not valid on various classes of frames. For example, for any formulas φ and ψ, we have:

- $\models \varphi \vee \neg\varphi$;
- $\models \Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$;
- $\models \Diamond(\varphi \wedge \psi) \to (\Diamond\varphi \wedge \Diamond\psi)$, but $\not\models (\Diamond\varphi \wedge \Diamond\psi) \to \Diamond(\varphi \wedge \psi)$;
- $\not\models \Box\varphi \to \varphi$, but $\mathcal{F}_{ref} \models \Box\varphi \to \varphi$;[17]
- $\mathcal{F}_{ref} \not\models \Box\varphi \to \Box\Box\varphi$, but $\mathcal{F}_{trans} \models \Box\varphi \to \Box\Box\varphi$.[18]

[17] $\mathcal{F}_{ref}$ denotes the class of all *reflexive* frames $F = (X, R)$, that is, where $(\forall x \in X)(xRx)$.

[18] $\mathcal{F}_{trans}$ denotes the class of all *transitive* frames $F = (X, R)$, that is, where $(\forall x, y, z \in X)((xRy \;\&\; yRz) \Rightarrow xRz)$.

Frame definability and invariance

The more we restrict a class of frames, the more formulas it might validate.[19] Given a formula φ, we might then try to find the *least restrictive* class of frames $\mathcal{F}$ that validates φ. This is formalized as in the following definition.

A formula φ **defines** a class of frames $\mathcal{F}$ if, for every frame F, $F \models \varphi$ if and only if $F \in \mathcal{F}$.[20] Some examples:

- Although $\mathcal{F}_{ref} \models p \to p$, this formula does not *define* $\mathcal{F}_{ref}$ since it is also valid on frames outside of this class;[21]
- $\Box p \to p$ defines the class $\mathcal{F}_{ref}$: this is because $\mathcal{F}_{ref} \models \Box p \to p$ and moreover, as we saw in class, any non-reflexive frame F can be equipped with a valuation function in such a way that the resulting model refutes $\Box p \to p$;
- $\Box p \to p$ does not define the class $\mathcal{F}_{id}$,[22] since even though $\mathcal{F}_{id} \models \Box p \to p$, one can find frames outside this class that also validate $\Box p \to p$;
- $\Box p \to \Box\Box p$ defines $\mathcal{F}_{trans}$: this is because $\mathcal{F}_{trans} \models \Box p \to \Box\Box p$ and moreover, as shown in class, any non-transitive frame F can be equipped with a valuation function in such a way that the resulting model refutes $\Box p \to \Box\Box p$.

There are many classes of frames that are *not* defined by any formula in the basic modal language. To prove such non-definability results, we make use of the concept of *invariance*.

Let x be a world in some model M. The **theory of** x **in** M is the set $Th_M(x) = \{\varphi : (M, x) \models \varphi\}$. In other words, it's the set of all formulas that are true at x. There can be different worlds that have the *same theory*—the idea of invariance is to investigate ways of recognizing when this happens.

Perhaps the simplest type of invariance is *invariance under disjoint union*. Given two models $M_1 = (X_1, R_1, v_1)$ and $M_2 = (X_2, R_2, v_2)$ with $X_1 \cap X_2 = \emptyset$, define the **disjoint union of** M_1 **and** M_2 to be the model $M_1 \sqcup M_2 = (X, R, v)$ where:

- $X = X_1 \cup X_2$;
- $R = R_1 \cup R_2$;
- $v(p) = v_1(p) \cup v_2(p)$.

Then one can show[23] that for all $x \in X_1$, $Th_{M_1}(x) = Th_{M_1 \sqcup M_2}(x)$.[24]

Another example of invariance is *invariance of generated submodels*. Given a model $M = (X, R, v)$ and a point $x \in X$, define the **generated submodel of** M **at** x to be the model $M_x = (R^*(x), R|_{R^*(x)}, v|_{R^*(x)})$,

[19] One way of thinking about this is that a more restricted class of frames has fewer counterexamples available.

[20] So the right-to-left direction is simply the requirement that $\mathcal{F} \models \varphi$, whereas the left-to-right requirement demands (in contrapositive) that whenever $F \notin \mathcal{F}$, some model based on F actually refutes φ.

[21] In fact, it is valid on *all* frames, so (trivially) it defines the class of all frames.

[22] That is, the class of all frames where R is the identity relation, so xRy iff $x = y$.

[23] We showed this in class.

[24] Of course, an analogous result holds for all $x \in X_2$.

where R^* is the reflexive, transitive closure of R.[25] Then it can be shown that for all $y \in R^*(x)$, $Th_{M_x}(y) = Th_M(y)$.

[25] Or, equivalently, $R^*(x)$ is the set of worlds R-reachable from x, where y is **R-reachable from** x if there exists a sequence $x_1, \dots, x_k$ such that $x_1 = x$, $x_n = y$, and for all $1 \leq i < n$, $x_i R x_{i+1}$.

Much more general than either of these notions of invariance is *invariance under bisimulation*. Given two models $M_1 = (X_1, R_1, v_1)$ and $M_2 = (X_2, R_2, v_2)$, a **bisimulation** between M_1 and M_2 is a binary relation $\sim \subseteq X_1 \times X_2$ such that, whenever $x_1 \sim x_2$, the following conditions are satisfied:

BASE for all $p \in$ PROP, $x_1 \in v_1(p) \Leftrightarrow x_2 \in v_2(p)$;

FORTH for all $y_1 \in R_1(x_1)$, there exists a $y_2 \in R_2(x_2)$ with $y_1 \sim y_2$;

BACK for all $y_2 \in R_2(x_2)$, there exists a $y_1 \in R_1(x_1)$ with $y_1 \sim y_2$.

Theorem 1. *If $\sim$ is a bisimulation between M_1 and M_2 and $x_1 \sim x_2$, then $Th_{M_1}(x_1) = Th_{M_2}(x_2)$.*

Proof. Induction on φ.[26] $\square$

[26] As shown in class.

As mentioned, invariance results are useful for proving that certain classes of frames cannot be defined. For example, the class $\mathcal{F}_{inf}$ of infinite frames cannot be defined. To see this, suppose for contradiction that φ defines $\mathcal{F}_{inf}$. Let F be any finite frame; then some model M based on F refutes φ, say at the point x. Now let M' be an infinite model disjoint from M. Consider $M \sqcup M'$; clearly this is an infinite model, so it should validate φ; however, by invariance under disjoint union, x refutes φ in $M \sqcup M'$ just as it does M. This contradiction shows that φ cannot define $\mathcal{F}_{inf}$.

We can also leverage bisimulations to help us prove frame nondefinability results. The key concept is the following: given two frames $F_1 = (X_1, R_1)$ and $F_2 = (X_2, R_2)$, a **bounded morphism** from F_1 to F_2 is a function $f : X_1 \to X_2$ that satisfies the FORTH and BACK conditions of a bisimulation.[27]

[27] Recall that every function is also a relation, so this makes sense. The BASE condition cannot be demanded since we are dealing with frames, not models.

Theorem 2. *If f is a surjective,[28] bounded morphism from F_1 to F_2, then if $F_2 \not\models \varphi$, also $F_1 \not\models \varphi$.*

[28] I.e., $(\forall y \in X_2)(\exists x \in X_1)(f(x) = y)$.

Proof. Since $F_2 \not\models \varphi$ there is a valuation v_2 such that $M_2 = (F_2, v_2) \not\models \varphi$. So some $y \in X_2$ is such that $(M_2, y) \not\models \varphi$. Define $v_1 :$ PROP $\to 2^{X_1}$ by $v_1(p) = f^{-1}(v_2(p))$. Then it is easy to check that f is a bisimulation between $M_1 = (F_1, v_1)$ and M_2. Since f is surjective, there is some $x \in X_1$ with $f(x) = y$. Then by Theorem 1, since $(M_2, y) \not\models \varphi$, also $(M_1, x) \not\models \varphi$, so $F_1 \not\models \varphi$, as desired. $\square$

Corollary 3. *If $\mathcal{F}$ is a class of frames, $F_1 \in \mathcal{F}$, and $F_2 \notin \mathcal{F}$, and there exists a surjective, bounded morphism from F_1 to F_2, then $\mathcal{F}$ is not definable by any formula in the basic modal language.*

Proof. Suppose for contradiction that φ defines $\mathcal{F}$. Then $F_2 \not\models \varphi$, so by Theorem 2, $F_1 \not\models \varphi$, a contradiction. $\square$

Finally, we can also use invariance results to show that certain ways of expanding the basic modal language genuinely increase our expressive power. For instance, although we might "extend" the basic modal language with the unary modality $\lozenge\varphi$ defined as follows:

$$x \models \lozenge\varphi \quad \text{iff} \quad (\exists y \in R(x))(y \models \varphi),$$

this does not increase the expressive power of the language because $\lozenge\varphi$ is equivalent to a formula in the basic modal langauge: it is easy to see that $\models \lozenge\varphi \leftrightarrow \neg\Box\neg\varphi$.[29]

On the other hand, we might extend the basic modal language with the unary modality $E\varphi$ defined as follows:[30]

$$x \models E\varphi \quad \text{iff} \quad (\exists y \in X)(y \models \varphi).$$

In this case we do obtain a genuinely more expressive language. To see this, consider two disjoint models M_1 and M_2 and suppose that $M_1 \models \neg p$ and $M_2 \not\models \neg p$. Let x be a point in M_1; then clearly $(M_1, x) \not\models Ep$, since $[\![p]\!]_{M_1} = \emptyset$. However, since $[\![p]\!]_{M_2} \neq \emptyset$, we can see that $(M_1 \sqcup M_2, x) \models Ep$. Invariance under disjoint union tells us that $Th_{M_1}(x) = Th_{M_1 \sqcup M_2}(x)$, so it follows that Ep cannot be equivalent to any formula in the basic modal language.

[29] Indeed, we previously defined $\lozenge\varphi$ to be an abbreviation for $\neg\Box\neg\varphi$.

[30] That is, $E\varphi$ is a "true" existential quantifier, ranging over the entire set of worlds, not just the worlds accessible from x.

Soundness and Completeness

A CORE FOCUS OF LOGIC is *deduction*—the idea of mechanically deriving new formulas from previously accepted formulas using some pre-defined rules. One common way to organize this is as follows: you start with some formulas called *axioms*, which are accepted without justification. Then, starting from these axioms, we apply *rules of inference* (which are also accepted without justification) to derive more formulas.

One of the standard axiom systems in modal logic is called K, and it specifies the following axioms and rules of inference:

- every tautology of propositional logic is an axiom;
- every instance of the scheme $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$ is an axiom;[31]
- from φ and $\varphi \to \psi$ we may derive ψ;[32]
- from φ we may derive $\Box\varphi$.[33]

[31] "Distribution".

[32] "Modus ponens".

[33] "Necessitation".

A **deduction from** K is a finite sequence of formulas $\varphi_1, \ldots, \varphi_n$ such that, for each $1 \leq i \leq n$, either:

(1) φ_i is an axiom, or

(2) there exist $j, k < i$ such that $\varphi_k = \varphi_j \to \varphi_i$,[34] or

(3) there exists $j < k$ such that $\varphi_i = \Box\varphi_j$.[35]

In this case, the sequence $\varphi_1, \ldots, \varphi_n$ is said to be a *deduction of* φ_n *from* K.[36] We write $\vdash_{\mathsf{K}} \varphi$ to indicate that there exists a deduction of φ from K.

[34] In other words, φ_i follows from two previous "lines" of the deduction by modus ponens.

[35] I.e., φ_i follows from a previous line by necessitation.

[36] It is easy to see (and useful to observe) that if $\varphi_1, \ldots, \varphi_n$ is a deduction, so is any truncation $\varphi_1, \ldots, \varphi_m$, where $m \leq n$.

The *goal* of deduction is to arrive at *truths*—or, more precisely, validities. It turns out that K is a very good axiom system for the following two reasons:

- Everything deducible from K is valid;
- Everything that is valid is deducible from K.

These properties are called *soundness* and *completeness*, respectively. Soundness says that deductions never "go wrong"—they always lead to validities. Completeness is even more impressive, because it says that deductions are powerful enough to reach *all* the validities.

Theorem 4. *The axiom system* K *is sound with respect to the class of all frames:* $\vdash_{\mathsf{K}} \varphi$ *implies* $\models \varphi$.

Proof. By induction on the length of the deduction of φ, as shown in class. The idea is to first show that every axiom is valid, then show that validity is preserved by modus ponens and necessitation. □

More generally, an axiom system may be sound or complete with respect to a *class* of frames. Given an axiom system AX and a class of frames $\mathcal{F}$, we say that AX is **sound** with respect to $\mathcal{F}$ if $\vdash_{\mathsf{AX}} \varphi$ implies $\mathcal{F} \models \varphi$, and **complete** with respect to $\mathcal{F}$ if $\mathcal{F} \models \varphi$ implies $\vdash_{\mathsf{AX}} \varphi$.

It is not hard to see, for example, that the axiom system called T, which is obtained by adding all instances of $\Box\varphi \to \varphi$ as axioms to K, is sound with respect to the class of all reflexive frames.[37]

[37] The proof of Theorem 4 simply needs to be augmented with an argument that each of the newly introduced axioms is valid in $\mathcal{F}_{ref}$.

Soundness results are also useful for showing us which formulas are *not* deducible in a given axiom system. For example, $\Box p \to p$ is not deducible from K since, if it were, by Theorem 4 we would have $\models \Box p \to p$, which we have seen is not so.

The canonical model construction

Proving completeness for K is significantly more involved.[38] The technique we will use is called the *canonical model construction*: we will build a gigantic model that simultaneously refutes *every* non-theorem of K. This is sufficient for our proof, since the contrapositive of completeness says that $\not\vdash_{\mathsf{K}} \varphi$ implies $\not\models \varphi$.

[38] We will begin by proving completeness for K, then extend the technique to other axiom systems and other types of completeness results.

How can we build this "canonical" model? Intuitively, we are going to put one world into the model for *every possible way things might be*. The way to do this formally is by defining the notion of *maximally consistent sets*.

Consider a finite set of formulas $\Gamma = \{\varphi_1, \ldots, \varphi_n\}$. Then Γ is said to be AX**-consistent** if $\not\vdash_{\mathsf{AX}} \neg(\varphi_1 \wedge \cdots \wedge \varphi_n)$.[39] Note that this definition has nothing to do with truth or validity—it's cashed out entirely in terms of deduction.

[39] So a finite set of formulas is AX-consistent if you can't deduce the negation of their conjunction in AX.

An arbitrary set of formulas Γ (not necessarily finite) is called AX**-consistent** if every finite subset of Γ is AX-consistent.[40] The following lemma will be useful later.

[40] So consistency of an infinite set just means you can't deduce the negation of (the conjunction of) any finite subset.

Lemma 5. *Suppose that* AX *includes as axioms all tautologies of classical propositional logic as well as the rule of inference modus ponens.*[41] *Then if*

[41] Henceforth, we assume that all axiom systems under consideration have this property—namely, that they all extend the deduction system of classical propositional logic.

$\Gamma \cup \{\varphi\}$ *is* AX*-inconsistent, there is a finite subset* $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma$ *such that* $\vdash_{\mathsf{AX}} (\varphi_1 \wedge \ldots \wedge \varphi_n) \to \neg\varphi$.

Proof. Since $\Gamma \cup \{\varphi\}$ is AX-inconsistent, there is a finite subset $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma$ such that either $\vdash_{\mathsf{AX}} \neg(\varphi_1 \wedge \cdots \wedge \varphi_n)$ or $\vdash_{\mathsf{AX}} \neg(\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \varphi)$. By propositional reasoning,[42] $\vdash_{\mathsf{AX}} \neg(\varphi_1 \wedge \cdots \wedge \varphi_n)$ implies $\vdash_{\mathsf{AX}} \neg(\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \varphi)$.[43] Once again using propositional reasoning,[44] this is equivalent to $\vdash_{\mathsf{AX}} (\varphi_1 \wedge \cdots \wedge \varphi_n) \to \neg\varphi$, as desired. $\square$

[42] Which applies because AX extends the deduction system of classical propositional logic.

[43] Apply deMorgan's law and observe that this is just an instance of disjunction introduction.

[44] Essentially the definition of the material conditional.

Finally, a set of formulas Γ is called **maximally AX-consistent** if it is AX-consistent and every proper superset of Γ is *not* AX-consistent.[45]

[45] In other words, a maximally AX-consistent set is a set of formulas that's "as big as it can be" while still remaining consistent.

Lemma 6. *Suppose that* Γ *is* AX*-consistent. Then* Γ *is maximally* AX*-consistent if and only if for every formula* φ*, either* $\varphi \in \Gamma$ *or* $\neg\varphi \in \Gamma$*.*

Proof. First suppose that for all φ, either $\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$. To show that Γ is maximally AX-consistent it suffices to show that $\Gamma' \supsetneq \Gamma$ implies Γ' is not AX-consistent. And indeed, if $\Gamma' \supsetneq \Gamma$, then there is some $\psi \in \Gamma' \setminus \Gamma$, but then, since $\psi \notin \Gamma$, we must have $\neg\psi \in \Gamma$, and so $\psi, \neg\psi \in \Gamma'$, from which it follows that Γ' is not AX-consistent.[46]

[46] Since $\vdash_{\mathsf{AX}} \neg(\psi \wedge \neg\psi)$.

Conversely, suppose that Γ' is maximally AX-consistent, and suppose for contradiction that $\varphi \notin \Gamma$ and also $\neg\varphi \notin \Gamma$. By maximality, it must then be that both $\Gamma \cup \{\varphi\}$ and $\Gamma \cup \{\neg\varphi\}$ are AX-inconsistent. By Lemma 5 this means that there is some finite subset $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma$ such that $\vdash_{\mathsf{AX}} (\varphi_1 \wedge \cdots \wedge \varphi_n) \to \neg\varphi$, and also some finite subset $\{\psi_1, \ldots, \psi_m\} \subseteq \Gamma$ such that $\vdash_{\mathsf{AX}} (\psi_1 \wedge \cdots \wedge \psi_m) \to \neg\neg\varphi$. But this implies that $\vdash_{\mathsf{AX}} (\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \psi_1 \wedge \cdots \wedge \psi_m) \to (\neg\varphi \wedge \neg\neg\varphi)$, which in turn implies that $\vdash_{\mathsf{AX}} \neg(\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \psi_1 \wedge \cdots \wedge \psi_m)$, contradicting the assumption that Γ is AX-consistent. $\square$

In light of this result, we might think of maximally AX-consistent sets as consisting of a long list of consistent "opinions" about the truth or falsity of *every single formula* in the basic modal language.

After all these definitions and all this work work, do we even know whether maximally AX-consistent sets exist? The answer is yes, they do, as follows from the following famous lemma.

Lemma 7 (Lindenbaum's Lemma)**.** *Let* Γ *be any* AX*-consistent set of formulas. Then there exists a maximally* AX*-consistent set* $\Gamma' \supseteq \Gamma$*.*

Proof. Let $\varphi_1, \varphi_2, \ldots$ be an enumeration of all formulas of the basic modal language. Set $\Gamma^{(0)} = \Gamma$, and inductively define[47]

[47] So we start with Γ and either add φ_1 or not, depending on whether it's consistent. Then similarly with φ_2, and φ_3, and so on.

$$\Gamma^{(k+1)} = \begin{cases} \Gamma^{(k)} \cup \{\varphi_{k+1}\} & \text{if this set is AX-consistent} \\ \Gamma^{(k)} & \text{otherwise.} \end{cases}$$

Define $\Gamma' = \bigcup_{k=0}^{\infty} \Gamma^{(k)}$. Clearly $\Gamma' \supseteq \Gamma$. Moreover, Γ' must be AX-consistent, since if it weren't some finite subset would already be inconsistent, but that finite subset would be contained in some $\Gamma^{(k)}$, contradicting the fact that each such set is AX-consistent. Finally, Γ' must be maximally AX-consistent, since for each formula φ, either it is in Γ' already, or else it was rejected for addition to some $\Gamma^{(k)}$ because $\Gamma^{(k)} \cup \{\varphi\}$ is AX-inconsistent, which of course implies that $\Gamma' \cup \{\varphi\}$ is also AX-inconsistent. $\square$

Corollary 8. *A maximally consistent set exists.*

Proof. By Lindenbaun's Lemma, since $\emptyset$ is AX-consistent, it can be extended to some maximally AX-consistent set Γ'. $\square$

At long last, we can build our canonical model $\tilde{M} = (\tilde{X}, \tilde{R}, \tilde{v})$. The set of worlds is given by

$$\tilde{X} = \{x_\Gamma \,:\, \Gamma \text{ is maximally AX-consistent}\}.$$

In other words, there is one world for every maximally AX-consistent set.[48] For any formula φ, define

$$\hat{\varphi} = \{x_\Gamma \in \tilde{X} \,:\, \varphi \in \Gamma\}.$$

So $\hat{\varphi}$ collects the set of worlds (in the canonical model) that correspond to maximally AX-consistent sets containing φ. Given that our worlds have a *built-in* connection to formulas, it is natural to define the canonical valuation $\tilde{v} : \text{PROP} \to 2^{\tilde{X}}$ by

$$\tilde{v}(p) = \hat{p}.$$

So we are saying that p is *true* at exactly those worlds that correspond to maximally AX-consistent sets that contain p. Finally, we define the canonical relation $\tilde{R} \subseteq \tilde{X} \times \tilde{X}$ as follows:

$$x_\Gamma \tilde{R} x_\Delta \text{ iff } (\forall \varphi)(\Box\varphi \in \Gamma \Rightarrow \varphi \in \Delta).$$

Intuitively: whenever Γ has the "opinion" that $\Box\varphi$, we are insisting that the only accessible worlds from x_Γ be worlds x_Δ where Δ has the "opinion" that φ.

The main point of all these definitions is to be able to prove the following lemma, which says that the "opinions" of each world in the canonical model match exactly with the formulas that are actually true at those worlds.

Lemma 9 (Truth Lemma)**.** *For all formulas φ, we have $[\![\varphi]\!]_{\tilde{M}} = \hat{\varphi}$.*[49]

Once we've proved this, we can establish the main result!

Corollary 10. K *is complete with respect to the class of all models.*

[48] This connects to the intuition, raised earlier, that a maximally AX-consistent set has an "opinion" about every formula—in total, it represents a full description (in the basic modal language) of one way the world might be.

[49] In other words, for all formulas φ and every $x_\Gamma \in \tilde{X}$, we have $(\tilde{M}, x_\Gamma) \models \varphi$ iff $\varphi \in \Gamma$.

Proof. It suffices to show that for every φ that is not deducible from K, we have $\tilde{M} \not\models \varphi$. And indeed, if $\nvdash_{\mathsf{K}} \varphi$, then clearly $\{\neg\varphi\}$ is K-consistent. Thus by Lindenbaum's Lemma there is a maximally K-consistent set $\Gamma' \supseteq \{\neg\varphi\}$. But then by the Truth Lemma, we have $x_{\Gamma'} \models \neg\varphi$, so $\tilde{M} \not\models \varphi$, as desired. □

So what remains is to prove the Truth Lemma. To do so, it will be useful to first establish the following result.

Lemma 11 (Existence Lemma)**.** *Suppose that* AX *contains all instances of the distribution axiom scheme, as well as the necessitation rule of inference. Let* Γ *be a maximally* AX*-consistent set, and suppose that* $\Diamond\varphi \in \Gamma$*. Then there exists a maximally* AX*-consistent set* Δ *such that* $\varphi \in \Delta$ *and* $x_\Gamma \tilde{R} x_\Delta$*.*

Proof. Since we are trying to show that a certain maximally AX-consistent set exists, it makes sense to try to use Lindenbaum's Lemma, somehow. We want Δ to have two properties: first, to contain φ, and second, to be accessible from Γ.[50] To be accessible from Γ, by definition, Δ must contain the formula ψ whenever Γ contains $\Box\psi$. So let's define the set

$$\Gamma^{\Box} = \{\psi : \Box\psi \in \Gamma\},$$

and observe that if we can prove the existence of maximally AX-consistent set Δ that contains $\Gamma^{\Box} \cup \{\varphi\}$, we'll be done!

This is where Lindenbaum's Lemma comes in: we'll show that $\Gamma^{\Box} \cup \{\varphi\}$ is AX-consistent, from which it follows that it can be extended to a maximally AX-consistent set, producing the desired set Δ. So suppose for contradiction that $\Gamma^{\Box} \cup \{\varphi\}$ is AX-inconsistent; then, by Lemma 5, we know that there is some finite collection of formulas $\psi_1, \ldots, \psi_n \in \Gamma^{\Box}$ such that

$$\vdash_{\mathsf{AX}} (\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \neg\varphi.$$

Then, by necessitation, we must also have

$$\vdash_{\mathsf{AX}} \Box((\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \neg\varphi),$$

and so by distribution and modus ponens we can deduce that

$$\vdash_{\mathsf{AX}} \Box(\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \Box\neg\varphi. \tag{1}$$

One can also show that[51]

$$\vdash_{\mathsf{AX}} (\Box\psi_1 \wedge \cdots \wedge \Box\psi_n) \rightarrow \Box(\psi_1 \wedge \cdots \wedge \psi_n),$$

so putting this together with (1) yields

$$\vdash_{\mathsf{AX}} (\Box\psi_1 \wedge \cdots \wedge \Box\psi_n) \rightarrow \Box\neg\varphi.$$

This is useful because the formulas $\Box\psi_1, \ldots, \Box\psi_n$ are all in Γ, and so it follows that $\Box\neg\varphi$ must also be in Γ.[52] But, by assumption, Γ also contains $\Diamond\varphi$, i.e., $\neg\Box\neg\varphi$, so this contradicts AX-consistency of Γ. □

[50] Or, more precisely, for x_Δ to be accessible from x_Γ; henceforth, I will stop harping on this minor distinction.

[51] We saw the case for $n = 2$ in class; the general result is left as an exercise to the reader.

[52] See Exercise 1.

Exercise 1

We used the following fact in the proof of the previous lemma: if Γ is maximally AX-consistent, $\chi_1, \dots, \chi_k \in \Gamma$, and

$$\vdash_{\mathsf{AX}} (\chi_1 \wedge \cdots \wedge \chi_n) \to \chi,$$

then $\chi \in \Gamma$. Prove this.

The Existence Lemma basically tells us that when Γ has the "opinion" that $\Diamond\varphi$, that opinion is backed up in the canonical model by the existence of an accessible-from-Γ world that has the opinion φ. This is crucial in the proof of the Truth Lemma.

Proof of the Truth Lemma. The proof proceeds by induction on φ. For the base case where $\varphi = p \in \textsc{prop}$, we have

$$[\![p]\!]_{\tilde{M}} = \tilde{v}(p) = \hat{p},$$

by definition of $\tilde{v}$.

So suppose inductively that the result holds for φ and ψ; we wish to show it holds for $\neg\varphi$, $\varphi \wedge \psi$, and $\Box\varphi$. Observe that

$$\begin{aligned} x_\Gamma \models \neg\varphi \quad &\Leftrightarrow \quad x_\Gamma \not\models \varphi \\ &\Leftrightarrow \quad \varphi \notin \Gamma \\ &\Leftrightarrow \quad \neg\varphi \in \Gamma. \end{aligned}$$

The first equivalence is simply the semantic definition of negation, the second is the inductive hypothesis, and the third follows from the fact that Γ is a maximally AX-consistent set, as the reader is invited to check.[53] The inductive step for conjunction is similar.

Now we attend to the hard part, the inductive step for $\Box$. First suppose that $\Box\varphi \in \Gamma$; we want to show that $x_\Gamma \models \Box\varphi$. Let $x_\Delta \in \tilde{R}(x_\Gamma)$; then by definition of $\tilde{R}$, we know that $\varphi \in \Delta$. But then the inductive hypothesis tells us that $x_\Delta \models \varphi$; since this holds for arbitrary $x_\Delta \in \tilde{R}(x_\Gamma)$, we have established that $x_\Gamma \models \Box\varphi$, as desired.

Conversely, suppose that $\Box\varphi \notin \Gamma$; we wish to show that $x_\Gamma \not\models \Box\varphi$. By maximality, we must have $\neg\Box\varphi \in \Gamma$, and thus $\Diamond\neg\varphi \in \Gamma$. Now by the Existence Lemma, we know there is some $x_\Delta \in \tilde{R}(x_\Gamma)$ with $\neg\varphi \in \Delta$. This implies that $\varphi \notin \Delta$, so by the inductive hypothesis, $x_\Delta \not\models \varphi$, and therefore $x_\Gamma \not\models \Box\varphi$. $\square$

[53] The forward implication comes from maximality, the backward implication from AX-consistency.

That's it! That's enough to establish completeness for K with respect to the class of all frames, as shown in Corollary 10.

Extending completeness results

As with the soundness result, we can extend this proof technique to establish completeness for other axiom systems as well. For example, let's suppose we want to prove that T is not only sound with respect to the class of all reflexive frames, but also complete with respect to this class. What should we do?

We can, once again, build a canonical model. In fact, the entirety of the canonical model construction above was carried out with respect to an arbitrary axiom system AX.[54] So it's a bit misleading to be calling it *the* canonical model—we actually built a whole family of canonical models, parameterized by the underlying axiom system.

[54] Well, not totally arbitrary: we assumed AX extends the deductive system of classical propositional logic, and moreover, to prove that Existence Lemma, we additionally assumed that AX is at least as strong as K.

What difference does the axiom system make? It determines the notion of consistency, which in turn determines which worlds there are! For example, the canonical model construction for K and for T contain different worlds, because there are sets that are maximally K-consistent but not maximally T-consistent. Indeed, one can prove the following.

Exercise 2

Any set that contains both $\Box p$ and $\neg p$ is T-inconsistent, yet there is a maximally K-consistent set that contains both these formulas.

So the canonical model construction produces a different model for T than it does for K, but all the lemmas we proved were independent of whether AX was equal to K or T or some other system. So are we done? Do we have completeness for T with respect to the class of reflexive frames?

The answer is no. What is missing? Try to figure it out before reading further...[55]

[55] Hint: ask yourself, where did the class of *reflexive* frames come into play?

If we look again at the proof of Corollary 10, we see that all we needed to show was that any formula not deducible from K is refuted on *some* model—it didn't matter which model. But now we are trying to prove that T is complete *with respect to the class of reflexive frames*—and for this we must show that any formula not deducible from T is refuted on some *reflexive* model. In other words, the completeness proof for T doesn't work unless the canonical model for T is reflexive.

Proposition 12. *The canonical model for* T *is reflexive.*

Proof. We need to show that for all x_Γ, $x_\Gamma \tilde{R} x_\Gamma$. By definition of $\tilde{R}$, this is equivalent to showing that whenever $\Box\psi \in \Gamma$, also $\psi \in \Gamma$. This follows from Exercise 1, since $\vdash_{\mathsf{T}} \Box\psi \to \psi$. $\Box$

This is a fairly general technique for using the canonical model construction to establish completeness of other axiom systems with respect to specific classes of spaces. You'll get lots of practice with it on the homework.

But this is not the only way to extend completeness results. We can also use the invariance techniques we studied previously to obtain more powerful or simply different completeness results for familiar axiom systems.

For a very simple example: K is sound and complete with respect to the class of all frames that contain a self-loop.[56] We have previously seen that K is sound and complete with respect to the class of *all* frames. Clearly this immediately implies that it is sound with respect to any smaller class of frames. But when we shrink the class of frames, completeness can fail—for instance, K is definitely *not* complete with respect to the class of reflexive frames, since there are non-theorems of K that cannot be refuted on this class.[57]

[56] This isn't a particularly useful thing to know, but it does serve as a good illustration of how completeness results can be strengthened.

[57] For instance, $\Box p \to p$.

Nonetheless, although shrinking a class of frames *can* ruin a completeness result, interestingly, there are many cases where it does not. This is an important conceptual distinction between "soundness and completeness" results and the "frame definability" results we studied previously. The fact that, for example, $\Box p \to p$ defines the class of reflexive frames *implies* that it does not define any other class of frames. By contrast, the fact that T is sound and complete with respect to the class of reflexive frames does *not* mean that it cannot be sound and complete with respect to some other class of frames as well. Indeed, we will see that it is.

Above we claimed that K is complete with respect to the class of frames that contain a self-loop. The proof is trivial.[58] Consider the model $M = (X, R, v)$ where $X = \{a\}$ and $R = \{(a, a)\}$.[59] Clearly M contains a self-loop. Let φ be a non-theorem of K, i.e., suppose that $\nvdash_{\mathsf{K}} \varphi$. Let $\tilde{M}$ be the canonical model for K. We know that $\tilde{M} \not\models \varphi$. By invariance under disjoint union, then, we also have $\tilde{M} \sqcup M \not\models \varphi$. Clearly $\tilde{M} \sqcup M$ contains a self-loop. Thus, we have shown that every non-theorem of K can be refuted on some model that contains a self-loop, which establishes the desired completeness result.

[58] Based on the previous examples we've seen involving the property of "containing a self-loop", it's reasonable to guess that the proof involves invariance under disjoint union in some way.

[59] It doesn't matter how v is defined.

This particular strengthening of the completeness result for K may not seem especially interesting, but we can do better. For example, a frame is called **connected** if it cannot be written as a disjoint union of two other frames.[60]

[60] Equivalently, a frame (X, R) is **connected** if for all distinct points $x, y \in X$, there is an $R^{\leftrightarrow}$-path from x to y, where $R^{\leftrightarrow}$ denotes the symmetric closure of R.

Exercise 3

Prove that K is sound and complete with respect to the class of connected frames.

One way of interpreting this result is as follows: we could have started this course by defining models in such a way as to *insist* that they are connected—i.e., ruling out disconnected models from the outset—and if we had done so, it wouldn't have changed our deductive system in any way. It would still be K. Succinctly: K cannot "tell" whether our models are connected.

One seemingly important aspect of our models is the fact that the relation is allowed to contain cycles.[61] But, as it turns out, we could've ruled that out too without changing the deductive system. In fact, we can push this even farther: we could've assumed that all frames are *trees*.

A frame (X, R) is called **rooted** if there exists a point $r \in X$ such that, for all $x \in X$, rR^*x.[62] In this case, r is called the *root*. A frame is called a **tree** if it is rooted with root r and moreover, for every $x \in X$, there is a *unique* R-path from r to x.

[61] A **cycle** is a finite collection of points $x_1, \ldots, x_k \in X$ such that for all $1 \leq i < k$, $x_i R x_{i+1}$, and also $x_k R x_1$.

[62] Recall that R^* denotes the reflexive, transitive closure of R.

Exercise 4

Show that if (X, R) is a tree then it cannot contain a cycle.

Exercise 5

Show that if (X, R) is a tree then for every $x \neq r$, x has a unique predecessor.

Proposition 13. K *is sound and complete with respect to the class of trees.*

Proof. Soundness is immediate. To establish completeness, we first define a transformation called *unravelling*: the idea is to start with an arbitrary rooted frame (not necessarily a tree), and "unravel" it about the root in a way that produces a tree (with the same root) that can refute the same formulas as the original frame.

How can we define such a transformation? Intuitively, whenever we can get to the same point in two different ways from r, we want to "count" it as a different point. The way to formalize this is to let the *R-paths* in F (starting at r) become the *worlds* of the unravelled frame. Given a rooted frame $F = (X, R)$ with root r, define

$$\vec{X} = \{(r, x_1, \ldots, x_k) \,:\, (r, x_1, \ldots, x_k) \text{ is an } R\text{-path in } F\},$$

and, given any two paths $\vec{x}, \vec{y} \in \vec{X}$, set

$$\vec{x}\vec{R}\vec{y} \text{ iff } \vec{x} = (r, x_1, \ldots, x_k) \text{ and } \vec{y} = (r, x_1, \ldots, x_k, y).$$

Exercise 6

Prove that $\vec{F} = (\vec{X}, \vec{R})$ is a tree with root (r) (that is, the root is the unique one-world path $(r) \in \vec{X}$).

The intuition here is that we have forced $\vec{F}$ to be a tree by encoding into each point $\vec{x} \in \vec{X}$ the unique $\vec{R}$-path from (r) to $\vec{x}$.[63]

Now consider the function $f : \vec{X} \to X$ defined by

$$f(r, x_1, \ldots, x_k) = x_k.$$

So f maps each path in $\vec{X}$ to the world (in F) where it ends. Note that this function is surjective, because F is rooted with root r. In fact, it is also a bounded morphism. For the forth condition, we need merely observe that $\vec{x}\vec{R}\vec{y}$ clearly implies $f(\vec{x})Rf(\vec{y})$. For the back condition, consider $f(r, x_1, \ldots, x_k) = x_k$ and suppose that $x_k R y$; then it is easy to see that $(r, x_1, \ldots, x_k)\vec{R}(r, x_1, \ldots, x_k, y)$, and of course $f(r, x_1, \ldots, x_k, y) = y$.

Since bounded morphisms pull back refutations, what we have shown is that any formula that can be refuted on a rooted frame can also be refuted on a tree! This is a powerful result; indeed, it allows us to immediate establish the desired result. Suppose $\not\vdash_{\mathsf{K}} \varphi$. Then φ is refuted on some frame, F, say at the point x. Then, by invariance under generated submodels, we know that φ can also be refuted on the generated subframe F_x. But F_x is rooted![64] So that means that φ can also be refuted on $\vec{F}_x$. Thus we have shown that every non-theorem of K can be refuted on a tree. $\square$

[63] Make sure you can do this exercise before you try to understand the rest of the proof!

[64] Indeed, it's root is x.

Corollary 14. K *is sound and complete with respect to the class of frames with no cycles.*

Proof. Apply Exercise 4. $\square$

Filtration and decidability

The next strengthening of completeness we will tackle is, arguably, even more impressive and useful. We will develop a technique called *filtration* that we will apply to show that K is complete with respect to the class of *finite* frames. This, in turn, will allow us to prove decidability of the deductive system.

The idea here is to take an arbitrary model and force it to be finite by "squishing" some its worlds together. To be sure, this kind of squishing will typically change the truth values of some formulas. But we can squish things "carefully" so that some formulas of interest have their particular truth values kept invariant by the squishing.

So let's start with a set Σ of formulas—intuitively, the ones we want to preserve the truth values of. It will be convenient to assume that Σ is **subformula-closed**, which simply means that whenever $\varphi \in \Sigma$ and ψ is a subformula of φ, also $\psi \in \Sigma$.

Now suppose we are given a model $M = (X, R, v)$. For each $x \in X$, let

$$|x|_\Sigma = \{y \in X \,:\, (\forall \varphi \in \Sigma)(y \models \varphi \Leftrightarrow x \models \varphi)\}.$$

It is easy to see that the collection $\{|x|_\Sigma \,:\, x \in X\}$ is a partition of X.[65] Intuitively, $|x|_\Sigma$ consists of all the worlds in X that agree with x about the truth value of every formula in Σ. This represents our "squishing": we are squishing together all the worlds that agree on the formulas in Σ.

[65] Recall that a *partition of* X is a collection of mutually disjoint sets that collectively cover X.

The **filtration of** M **through** Σ is the model $M_\Sigma = (X_\Sigma, R_\Sigma, v_\Sigma)$ defined as follows:

- $X_\Sigma = \{|x|_\Sigma \,:\, x \in X\}$;
- $|x|_\Sigma R_\Sigma |y|_\Sigma$ iff for some $x' \in |x|_\Sigma$ and $y' \in |y|_\Sigma$ we have xRy;
- $v_\Sigma(p) = \begin{cases} \{|x|_\Sigma \,:\, (\forall x' \in |x|_\Sigma)(x' \in v(p))\} & \text{if } p \in \Sigma \\ \emptyset & \text{otherwise.} \end{cases}$

Note that when $p \in \Sigma$, all the worlds in $|x|_\Sigma$ agree on the truth value of p, so the universal quantifier in the definition of v_Σ could just as well have been an existential quantifier.[66]

[66] The fact that $v_\Sigma(p) = \emptyset$ when $p \notin \Sigma$ is entirely arbitrary—we simply don't care about how formulas outside of Σ behave.

The key property of M_Σ is that it preserves the truth values of the formulas in Σ.

Proposition 15. *For every φ, if $\varphi \in \Sigma$ then for all $x \in X$, we have $(M, x) \models \varphi$ iff $(M_\Sigma, |x|_\Sigma) \models \varphi$.*

Proof. Induction on φ. For the base case, suppose $\varphi = p \in \textsc{prop}$ and $p \in \Sigma$. Then we have

$$(M, x) \models p \Leftrightarrow x \in v(p) \Leftrightarrow |x|_\Sigma \in v_\Sigma(p) \Leftrightarrow (M_\Sigma, |x|_\Sigma) \models p,$$

as desired.

The inductive steps for negation and conjunction are easy.[67] So suppose inductively that the result holds for φ; we wish to show it holds for $\Box\varphi$. We assume that $\Box\varphi \in \Sigma$, otherwise this is trivial.

[67] Though it is important that Σ is subformula-closed, here. Try it to see why.

First suppose that $(M, x) \models \Box\varphi$. We wish to show that $(M_\Sigma, |x|_\Sigma) \models \Box\varphi$. So let $|y|_\Sigma \in R_\Sigma(|x|_\Sigma)$ be an arbitrary world in M_Σ that is accessible from $|x|_\Sigma$. By definition, this means that there is some $x' \in |x|_\Sigma$ and some $y' \in |y|_\Sigma$ such that $x'Ry'$. Since $x' \in |x|_\Sigma$ and $\Box\varphi \in \Sigma$ we know that $x' \models \Box\varphi$. Thus $y' \models \varphi$. Thus, by the inductive hypothesis,[68] we know that $|y'|_\Sigma \models \varphi$. Since $y' \in |y|_\Sigma$, we have $|y|_\Sigma = |y'|_\Sigma$, so $|y|_\Sigma \models \varphi$, which establishes that $|x|_\Sigma \models \Box\varphi$, as desired.

Conversely, suppose that $(M_\Sigma, |x|_\Sigma) \models \Box\varphi$. We wish to show that $(M, x) \models \Box\varphi$. So let $y \in R(x)$. Since xRy, we know by definition that $|x|_\Sigma R_\Sigma |y|_\Sigma$, thus $|y|_\Sigma \models \varphi$. By the inductive hypothesis, then, we have $y \models \varphi$, which establishes that $x \models \Box\varphi$, as desired. □

[68] Here we use the fact that Σ is subformula-closed, so $\varphi \in \Sigma$.

So we have shown that filtrations preserve the truth values of all formulas in Σ. But what does this have to do with *finite* models? The connection lies in the following fact.

Proposition 16. *If Σ is finite, so is M_Σ. In fact, M_Σ cannot contain more than $2^{|\Sigma|}$ worlds.*[69]

[69] Where $|\Sigma|$ denotes the cardinality if Σ.

Exercise 7

Prove Proposition 16. *(Hint: First try it for $\Sigma = \{p\}$ to get a sense of why this is true.)*

At last, we are ready to prove the completeness theorem we have been aiming for.

Theorem 17. K *is complete with respect to the class of all finite frames.*

Proof. Suppose that $\nvdash_{\mathsf{K}} \varphi$. Then φ is refuted on some model, M. Let Σ_φ denote the set of all subformulas of φ. Clearly, Σ_φ is finite, therefore by Proposition 16, so is M_{Σ_φ}. It is also clear that Σ_φ is subformula-closed, so from Proposition 15 we can deduce that, since $M \not\models \varphi$, also $M_{\Sigma_\varphi} \not\models \varphi$. We have therefore shown that φ is refuted on a finite model. □

From this result we can also obtain *decidability* of the axiom system K. Roughly speaking, an axiom system is called decidable if there is an algorithmic method[70] to determine, for any given formula φ, whether or not $\vdash_{\mathsf{K}} \varphi$.[71]

[70] That is, something that could be implemented on a computer.

[71] We will not delve into the formal definition of computability here, but instead present these results at an intuitive level.

How could such an algorithm work? Since the definition of a deduction is completely finitistic, one thing we might imagine is a computer program that simply keeps trying different proofs. If we're clever, we can arrange it so that it eventually gets to any given proof, and therefore (eventually!) finds a proof for everything that has a proof.

This sounds good, but it's actually not enough. What we want is a decision procedure that says "yes" or "no" for every input φ, answering the question of whether it's deducible from K. But of course, the answer has to come in finite time! If we start with a φ that happens to actually be deducible from K, then the process described above will eventually find a proof and spit out "yes". However, if $\nvdash_{\mathsf{K}} \varphi$, then the process above will simply keep searching, forever. There will be no point at which we can be sure that it's definitively failed to find a proof—there's always the chance that the next proof it checks will be a proof of φ. So we can *never* output "no".

Exercise 8

What is wrong with the following argument: If φ is not deducible from K, then $\neg\varphi$ is, so we can simply wait to see which of φ or $\neg\varphi$ is proved.

But there is another way to design a decision algorithm, using the results of Theorem 17 and Proposition 16. Start with the formula φ, and count how many subformulas it has.[72] Denote this number by $n(\varphi)$. Proposition 16 then tells us that for any model M, M_{Σ_φ} contains at most $2^{n(\varphi)}$ worlds. Moreover, Theorem 17 tells us that $\nvdash_{\mathsf{K}} \varphi$ iff for some model M, $M_{\Sigma_\varphi} \not\models \varphi$. Therefore to check whether φ is deducible, it is enough to check each world in each model of size up to $2^{n(\varphi)}$—of which, of course, there are only finitely many.[73] Therefore this is something a computer can do. If φ is true at all worlds in all such models, then it must be deducible. If not, then it's not. Thus we have specified a finite algorithm that answers "yes" or "no", as desired.

[72] This is something a computer can easily do.

[73] Actually, this is a lie—there are infinitely many if we include all possible valuations. But valuations that agree on all the primitive propositions that occur in φ are effectively the same, for our purposes—so there are only finitely many we need to check.

Topological Semantics

TOPOLOGY IS THE MATHEMATICAL STUDY of space without distance,[74] what is sometimes called "rubber sheet geometry". Perhaps the most canonical image associated with the study of topology is that of a coffee cup deforming into a doughnut (and vice-versa).[75] What could this possibly have to do with modal logic?

[74] Whatever that means.

[75] These shapes are topologically equivalent.

In this chapter I will provide a somewhat idiosyncratic (though complete and rigorous) introduction to topology, and explain the connection to modal logic along the way. Perhaps the strongest intuitive connection surfaces through an *epistemic* interpretation of modality, so I will favour that metaphor as we go.[76]

[76] If you haven't seen topology before, don't worry—this is designed to be accessible to a newcomer. If you have seen topology before, you probably haven't seen it introduced in quite this way. So there's something new for everyone.

Topology

One way to approach topology is through a dissatisfaction with the concept of set membership. Set membership is a binary affair: given $x \in X$ and $A \subseteq X$, either $x \in A$ or $x \notin A$—there is no middle ground.

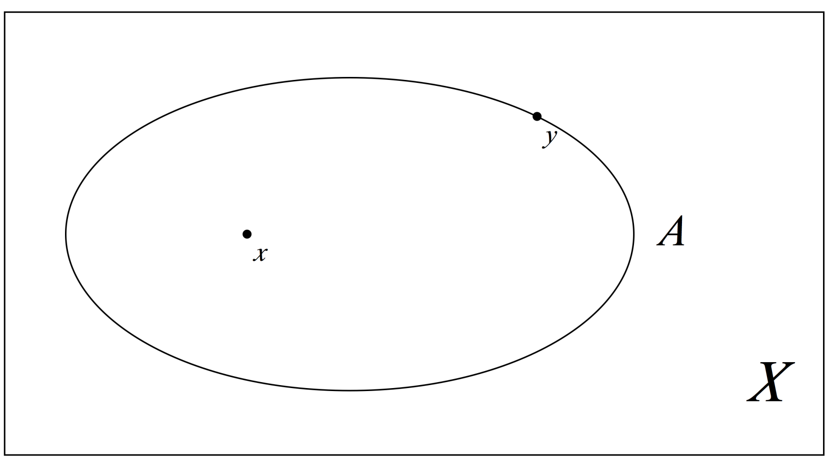

Figure 1: Set membership is binary.

However, drawing on very basic spatial intuitions, we can look at a diagram like the one in Figure 1 and see that x is "more" in A than y is. We might say x is "fully" or "robustly" in A, whereas y is just "barely" in A.[77]

How can we make sense of this? In other words, how could we (minimally) augment the basic structure of sets to represent this distinction between mere membership and "robust" membership?

One idea that immediately comes to mind is to incorporate a notion of distance. We can see that y, unlike x, is arbitrarily close to elements of $X \setminus A$. So if we added a distance function $d : X \times X \to [0, \infty)$, we could try to use it to capture this idea.

The topological approach is actually more subtle than that. Specifying a quantitative distance function[78] is overkill. We can represent spatial structure, and in particular capture a notion of "nearness", in purely qualitative terms—without a numeric distance function. This is what topology does.[79]

One intuition that might occur to us, if we stare at Figure 1 for a while, is that we if could pick up the set X and "shake" it a little, leaving everything just slightly displaced from its original location, then x would end up still being in A, whereas y might not. The same intuition can also be approached not with a "shaking" metaphor, but by thinking in terms of measurement and error.[80] If were were to try to measure the location of x, using (of course) our imperfect measurement tools, we would not get an exact, infinitely precise location, but rather a range of possible locations (represented by a set of points). Similarly for y. But as long as our measurement is precise enough, we can use it ascertain that x is indeed in A; see Figure 2.

[77] As usual, I adopt the convention of using a solid line to indicate a set that contains its boundary. So y really is in A.

[78] What is often called a *metric*.

[79] It's worth taking a moment, before reading further, to explore your own ideas about how one might capture the distinction between points like x and y in Figure 1 without referring to a quantitative notion of distance.

[80] This is the first point of contact with epistemology.

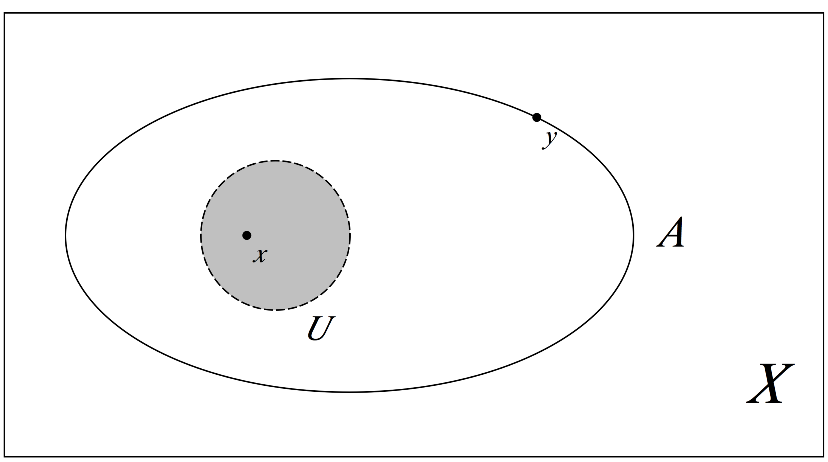

Figure 2: A witness to x's membership in A.

By contrast, intuitively, no measurement, no matter how precise, seems capable of telling us for sure that y is in A. See Figure 3.

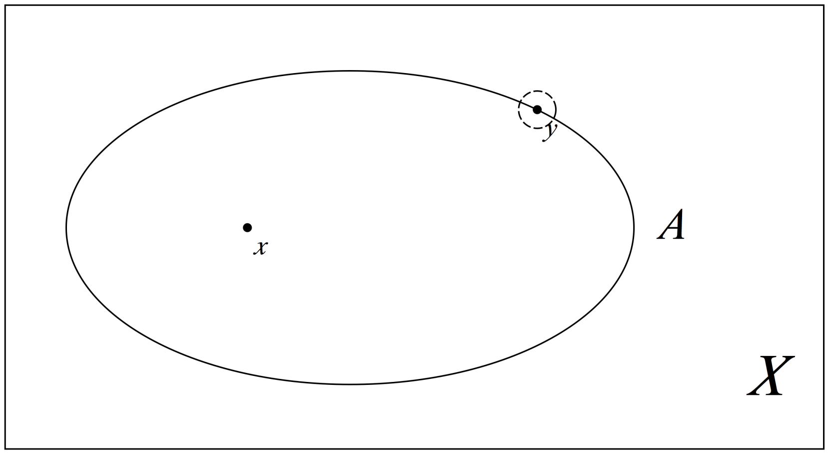

Figure 3: Trying to measure y's location.

This leads to the following first attempt to define the notion of "robust" membership: say that x is *robustly in A* if there exists a "witnessing" set U such that $x \in U$ and $U \subseteq A$. In other words, not only is x a member of A, but there is a kind of "cushion", in the form of the set U, that witnesses x's membership in A.

Once again, it's worth pausing here to ask yourself whether this is a reasonable definition. In fact, it's not. Can you see why?

Well, for one thing, we can always just take $U = \{x\}$, in which case we certainly have $x \in U$, and $U \subseteq A$ just in case $x \in A$. So in fact this definition of robust membership collapses just to plain old membership.

One might object that choosing a singleton set is "cheating", and so modify the definition to rule out singletons. But that doesn't really solve our problem either: we can just take $U = A$, in which case clearly $U \subseteq A$, and once again whenever $x \in A$ we have $x \in U$. So again we are left with a definition that doesn't actually make robust membership any different from regular membership.

The solution, and the core idea of topology, is to include the collection of all possible "witnessing" sets U as part of the fundamental structure of the space. That is, a **topological space** is a pair $(X, \mathcal{T})$ where X is a set, and $\mathcal{T} \subseteq 2^X$ is a collection of subsets of X satisfying certain properties.[81] The collection $\mathcal{T}$ is called a *topology* on X, and it's elements are called *open sets*.

[81] We'll come back to these properties later; for now they are a distraction.

Intuitively, open sets are precisely the sets that can act as "witnesses" for robust membership. This is formalized in the following definition: a point x is said to be in the **interior** of a set A if there exists an open set $U \in \mathcal{T}$ such that $x \in U \subseteq A$. In this case we write

$x \in int(A)$, so $int(A)$ denotes the *interior* of A—the set of all points in the interior of A.

So what we had been calling "robust membership" corresponds to the topological notion of interior, with the crucial addition of the topology, which encodes spatial structure by specifying which sets can act as witnesses.

If we return to the epistemic metaphor of measurement and error, the open sets correspond exactly to the *results of possible measurements or observations*, and, as such, $x \in int(A)$ can be interpreted as saying that there exists some measurement or observation that entails A, at x. Looking again at Figure 1, we see that while A is true at both x and y,[82] A is only *measurably* true at x.

[82] Saying "A is true at x" is just another way of saying that $x \in A$, borrowing terminology (and intuition) from possible world semantics.

It is easy to check that for all A, $int(A) \subseteq A$; that is, the interior function $int : 2^X \to 2^X$ is a *shrinking* operator—it maps every set to a subset of itself. This corresponds to the obvious fact that anything that's measurably true is also just true. It's also not hard to show that $int(int(A)) = int(A)$; that is, the interior operator is *idempotent*—applying it a second time does nothing.[83]

[83] Try to prove these facts yourself, before reading the proof below.

Proposition 18. *For all A, $int(A) \subseteq A$ and $int(int(A)) = int(A)$.*

Proof. First we'll show that $int(A) \subseteq A$. Let $x \in int(A)$. Then by definition there is a $U \in \mathcal{T}$ such that $x \in U$ and $U \subseteq A$, from which it follows immediately that $x \in A$, as desired.

Now we'll show that $int(A) \subseteq int(int(A))$ (the reverse containment is an immediate consequence of the previous result). Let $x \in int(A)$. Once again, by definition this means that there is a $U \in \mathcal{T}$ such that $x \in U \subseteq A$. We'll show that in fact $U \subseteq int(A)$, which will then yield (by definition) that $x \in int(int(A))$, as desired. So let $y \in U$. Then we have $y \in U \subseteq A$, so $y \in int(A)$; since y was arbitrary, this shows that $U \subseteq int(A)$ and completes the proof. $\square$

Exercise 9

Show that for all $A, B \subseteq X$, if $A \subseteq B$, then $int(A) \subseteq int(B)$.

The interior operator has other properties as well, but to prove them we need the full definition of a topological space.[84] Here it is: a **topological space** is a set X together with a collection $\mathcal{T} \subseteq 2^X$ such that

[84] Recall that we postponed a discussion of the properties that $\mathcal{T}$ must satisfy, previous.

(T1) $\mathcal{T}$ covers X;[85]

(T2) $U, V \in \mathcal{T} \Rightarrow U \cap V \in \mathcal{T}$;[86]

(T3) $\mathcal{C} \subseteq \mathcal{T} \Rightarrow \bigcup_{U \in \mathcal{C}} U \in \mathcal{T}$.[87]

[85] So for all $x \in X$, there is at least one $U \in \mathcal{T}$ with $x \in U$.

[86] $\mathcal{T}$ is closed under pairwise intersections.

[87] $\mathcal{T}$ is closed under arbitrary unions.

Property (T2) says that the intersection of any two open sets is open, while (T3) tells us that the union of *any* collection of open sets (no matter how big the collection is) is open.[88]

[88] Closure under pairwise intersection makes sense, from the epistemic perspective—it essentially corresponds to the idea that if we take two measurements or make two observations, we can "combine" what we learned from each one. Closure under arbitrary union, on the other hand, is a bit harder to interpret epistemologically. We'll return to this conceptual point later.

Exercise 10

Show that $\mathcal{T}$ is in fact closed under all finite intersections: that is, if $U_1, \dots, U_k \in \mathcal{T}$, then $\bigcap_{i=1}^k U_i \in \mathcal{T}$.

(T1) and (T3) together imply that $X \in \mathcal{T}$. We also always have $\emptyset \in \mathcal{T}$, because the empty set is the result of the "empty union".[89]

[89] That is, when $\mathcal{C} = \emptyset$, we have $\bigcup_{U \in \mathcal{C}} U = \emptyset$. It's probably best to think of this as a convention rather than a deep mathematical fact.

Exercise 11

Show that $int(X) = X$ and for all $A, B \subseteq X$, $int(A \cap B) = int(A) \cap int(B)$.

Perhaps at this point you've noticed an interesting connection.

Topology	Modal Logic
$int(A) \subseteq A$	$\Box\varphi \to \varphi$
$int(A) \subseteq int(int(A))$	$\Box\varphi \to \Box\Box\varphi$
$int(A \cap B) = int(A) \cap int(B)$	$\Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$
$int(X) = X$	$\Box(\varphi \vee \neg\varphi)$
if $A \subseteq B$ then $int(A) \subseteq int(B)$	from $\varphi \to \psi$ deduce $\Box\varphi \to \Box\psi$

Table 1: A comparison of topological properties and modal axioms and rules of inference.

On the left we have a list of several of the properties we've proved of topological spaces. And on the right, a list of axioms and a rule of inference that look suspiciously similar to the topological expressions on the left.[90]

[90] In fact, as you've seen on the problem sets, the righthand list comprises a sound and complete axiomatization of the class of reflexive and transitive frames.

The way to make this connection precise—and the cornerstone of the topological interpretation of modality—is to redefine the semantic interpretation of the box modality: instead of interpreting modal formulas on frames, we can interpret then on topological spaces;[91] and instead of defining $[\![\Box\varphi]\!]$ using a relation, we define it using the interior operator:

[91] Equipped, of course, with valuations.

$$[\![\Box\varphi]\!] = int([\![\varphi]\!]).$$

[NEW MATERIAL BEGINS HERE]

Before diving into the logic, however, it will be useful to look at a few concrete examples of topological spaces—if nothing else as an antidote to all the abstraction.[92] One easy class of examples comes from taking any set X and equipping it with the full powerset as a topology: $(X, 2^X)$. Trivially, 2^X satisfies all the requirements for a

[92] Of course, these may also be helpful on the problem set... But if you're already familiar with topology, feel free to skim this part.

topology. Any such space is called a *discrete space*. The geometric intuition is that in a discrete space, every point is "far" from every other point—for each point $x \in X$, there is a neighbourhood of x,[93] namely $\{x\}$, that contains no other points at all. Switching to the measurement analogy, an discrete space in one in which *every* fact is measurable—this corresponds to the fact that in such a space, for all $A \subseteq X$, $int(A) = A$.

[93] In a topological space, any open set containing x is called a *neighbourhood of* x.

At the other extreme, we can also form a topological space by starting with any set X and equipping it with the collection $\{\emptyset, X\}$ as a topology. Once again, it is easy to check that this collection satisfies all the requirements for a topology. These spaces are called *indiscrete*. Thinking geometrically, the image is of every point being squished together with every other point. Indeed, we have $int(A) = \emptyset$ whenever $A \neq X$—no point is robustly inside any set except for the whole space X itself. Thinking in terms of measurement, this says that no fact is measurable except those facts that are trivially true (in the sense of being true at every point).

Moving away from these extreme examples, but still thinking simple, if $X = \{a, b\}$ then the collection $\mathcal{T} = \{\emptyset, \{a\}, X\}$ is a topology on X.[94] One easily checks that $\mathcal{T}$ satisfies the requirements for a topology. It is a useful illustration of the fact that topological structure need not be "symmetric"—in the Sierpinski space we see that a is distinguishable from b (in the sense that there is a measurement that can be taken at a, namely $\{a\}$, that rules out b), but b is *not* distinguishable from a (the only measurement that can be taken at b is X itself, which is also compatible with a). We also have $int(\{a\}) = \{a\}$ but $int(\{b\}) = \emptyset$.

[94] This is the smallest example of a topological space that is neither discrete nor indiscrete. It is called the Sierpinski space.

One of the most important examples of a topological space is the set of real numbers, $\mathbb{R}$, equipped with its "usual" topology: $\mathcal{T}_{\mathbb{R}}$, which consists of all unions of open intervals.[95] This collection is closed under arbitrary unions by construction, but is it closed under finite intersections? The answer is yes, as follows from Proposition 19, below, but it is a useful exercise in the basic set theory of $\mathbb{R}$ to try to prove this directly.

[95] An open interval of $\mathbb{R}$ is any set of the form $(a, b) = \{x \in \mathbb{R} : a < x < b\}$, where $a < b$. Note that the inequalities here are strict! The word "open" is suggestive here, since each open interval will indeed become an open set in $\mathcal{T}_{\mathbb{R}}$.

Exercise 12

Prove that $\mathcal{T}_{\mathbb{R}}$ is closed under finite intersections.

This way of specifying a topology—as "all unions" of sets from some other collection—is important enough to articulate in generality. Given a set X, a collection $\mathcal{B} \subseteq 2^X$ is called a **basis for a topology** if

(B1) $\mathcal{B}$ covers X;

(B2) for all $U, V \in \mathcal{B}$ and any $x \in U \cap V$, there exists a set $W \in \mathcal{B}$

such that $x \in W \subseteq U \cap V$.

Exercise 13

Show that if $\mathcal{B}$ is closed under finite intersections, then (B2) is satisfied, but the reverse implication does not hold in general.

If $\mathcal{B}$ is a basis then $\mathcal{T}(\mathcal{B})$ is defined to be the collection of all unions of elements of $\mathcal{B}$, called the *topology generated by* $\mathcal{B}$. It is indeed a topology, as we now show.

Proposition 19. *If $\mathcal{B}$ is a basis for a topology, then $\mathcal{T}(\mathcal{B})$ is a topology.*

Proof. We need to show that $\mathcal{T}(\mathcal{B})$ satisfies (T1)–(T3). (T1) and (T3) are trivial, so we focus on (T2). Let $U, V \in \mathcal{T}(\mathcal{B})$; we wish to show that $U \cap V \in \mathcal{T}(\mathcal{B})$. Let $x \in U \cap V$. Since U and V are both unions of elements of $\mathcal{B}$, that means there exist $U_x, V_x \in \mathcal{B}$ such that $x \in U_x \subseteq U$ and $x \in V_x \subseteq V$. By (B2), then, there exists a set $W_x \in \mathcal{B}$ with $x \in W_x \subseteq U_x \cap V_x$. It follows that $W_x \subseteq U \cap V$ as well. Since we can find such a set for each $x \in U \cap V$, we can form the union

$$\bigcup_{x \in U \cap V} W_x.$$

This union clearly contains $U \cap V$, since it contains each $x \in U \cap V$; on the other hand, it is also contained in $U \cap V$, since each W_x is. Therefore

$$\bigcup_{x \in U \cap V} W_x = U \cap V.$$

This then shows that $U \cap V$ is a union of elements of $\mathcal{B}$, so by definition $U \cap V \in \mathcal{T}(\mathcal{B})$.[96] □

Getting back to the real line, we can understand the collection of open intervals to be a basis for the usual topology on $\mathbb{R}$.[97] This means, for example, that the interior of any *closed* interval[98] $[a, b]$ is exactly (a, b). To see this, first note that any point $x \in (a, b)$ is in $int([a, b])$ because (a, b) itself is a neighbourhood of x contained in $[a, b]$. On the other hand, any open neighbourhood U of a must contain an open interval (c, d) that contains a,[99] and therefore U must contain points less than a (and greater than c), which implies $U \not\subseteq [a, b]$. Similarly for b.

This reasoning is worth generalizing:

Proposition 20. *Let $(X, \mathcal{T})$ be a topological space, and let $\mathcal{B}$ be a basis for $\mathcal{T}$.*[100] *Then $x \in int(A)$ if and only if there is some $U \in \mathcal{B}$ such that $x \in U \subseteq A$.*

In other words, in the definition of the interior, it's okay to restrict attention to *basic open sets*.[101]

[96] Compare this proof to your solution to Exercise 12. How similar are they? Which one is easier?

[97] Similarly, we can view the collection of open balls as a basis for the usual topology on $\mathbb{R}^2$. (The open ball with radius ε centered at $(x, y) \in \mathbb{R}^2$ is defined to be the set

$$\{(x', y') \in \mathbb{R}^2 : \sqrt{(x - x')^2 + (y - y')^2} < \varepsilon\}.$$

That is, the set of all points whose Euclidean distance from (x, y) is (strictly) less than ε.)

[98] A closed interval of $\mathbb{R}$ is any set of the form $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$, where $a \leq b$. Note that the inequalities here are not strict! The word "closed" is suggestive here as well, as we shall soon see.

[99] Since U is a union of open intervals.

[100] That is, $\mathcal{T}(\mathcal{B}) = \mathcal{T}$.

[101] That is, open sets that occur in some basis for the topology.

Exercise 14

Prove Proposition 20.

To close this section, and in light of the previous example, we ought to say a word about *closed* sets. The easiest way to define a closed set is to say it's the complement of an open set. But this is perhaps not the best for developing intuition. For that we should instead start with the *closure* operator.

Similarly to the interior operator, we can also define the closure of a set—intuitively, the closure of A consists of all those points that are "almost" in A.[102] Once again, we use the topology to capture the idea of closeness, and say that x is in the **closure** of A if, for every open set $U \in \mathcal{T}$ that contains x, we have $U \cap A \neq \emptyset$. In other words, it's impossible to "separate" x from A with an open set.[103]

When x is in the closure of A we write $x \in cl(A)$, so $cl(A)$ is the set of all points in the closure of A. It's not hard to see that x is in the closure of A iff it is not in the interior of the complement of A, and similarly, x is in the interior of A iff it is not in the closure of the complement of A.[104]

Exercise 15

Prove that $cl(A) = X \setminus int(X \setminus A)$ and $int(A) = X \setminus cl(X \setminus A)$.

It's not hard to show that the interior of any set is open[105]—in fact, the open sets are precisely the fixed points of the interior operator.[106] Thus, we call a set **closed** if it is equal to its own closure; it can then be shown that a set is closed iff its complement is open.[107]

The logic of space

At last, we get to the logic part. A **topological model** is a topological space $(X, \mathcal{T})$ together with a valuation $v : \textsc{prop} \to 2^X$. Truth in a model $M = (X, \mathcal{T}, v)$ is defined via the following semantic clauses:

$$\begin{array}{rcl} [\![p]\!]_M & = & v(p) \\ [\![\neg\varphi]\!]_M & = & X \setminus [\![\varphi]\!]_M \\ [\![\varphi \wedge \psi]\!]_M & = & [\![\varphi]\!]_M \cap [\![\psi]\!]_M \\ [\![\Box\varphi]\!]_M & = & int([\![\varphi]\!]_M). \end{array}$$

In other words, the only difference is in the recursive clause for the box modality, as discussed. Based on the epistemic interpretation of the interior operator we've developed, we might naturally read

[102] Including points that are actually in A. Thus, the closure will be an expanding operator: mapping each set to a superset.

[103] In the measurement analogy: every measurement one might take at x is at least compatible with A; that is, A is *unfalsifiable*.

[104] Intuitively: x is "almost" in A iff it is not "robustly" outside of A, and x is "robustly" in A iff it is not almost outside of A.

[105] Consider the union of all open subsets of A.

[106] That is, $A \in \mathcal{T}$ iff $int(A) = A$.

[107] Compare Exercise 15.

$\Box\varphi$ as "φ is measurably true" or even "φ is knowable", though for the moment we will try to develop the mathematics as generally as possible.[108]

It's not hard to show that under these semantics, the axiom system S4[109] is sound with respect to the class of all topological spaces. The proof is quite similar to proofs of soundness in relational semantics, and is based primarily on the properties of topological spaces summarized in Table 1.

[108] That is, without pre-committing ourselves to a specific interpretation of $\Box$. In this spirit, a more "neutral" reading of $\Box\varphi$ might be "φ is robustly true", where the concept of "robustness" is itself left unspecified and therefore open to multiple interpretations.

[109] S4 is the axiom system that extends T with the axiom scheme $\Box\varphi \to \Box\Box\varphi$.

Exercise 16

Prove that S4 is sound with respect to the class of all topological spaces. [This is already a homework exercise.]

It turns out that S4 is also *complete* with respect to the class of all topological spaces.[110] To show this, we might try to construct a "canonical topological space"—and indeed this can be done and is quite illustrative. But we can actually prove this by making use of a completeness result we already know from relational semantics, namely, that S4 is complete with respect to the class of all reflexive and transitive frames.

[110] This is sometimes expressed by saying that "S4 is the logic of space".

Here's the idea: we'll show how to transform any reflexive and transitive frame into a topological space in a way that preserves the truth values of formulas. This will then show that anything that can be refuted in a reflexive and transitive frame can be refuted in a topological space, from which the desired completeness result follows.

Proposition 21. *Let (X, R) be a reflexive and transitive frame and define $\mathcal{B}_R = \{R(x) : x \in X\}$. Then $\mathcal{B}_R$ is a basis for a topology on X.*

Proof. First we need to show that $\mathcal{B}_R$ covers X. But this is easy: for each $x \in X$, by reflexivity, we have $x \in R(x)$, and by definition, $R(x) \in \mathcal{B}_R$.

Next, let $R(x), R(y) \in \mathcal{B}_R$ and take $z \in R(x) \cap R(y)$. We will show that $R(z) \subseteq R(x) \cap R(y)$, which will establish (B2). Consider any $z' \in R(z)$. We know that xRz (since $z \in R(x)$) and zRz' (since $z' \in R(z)$); thus, by transitivity, we have xRz'. This shows that $z' \in R(x)$. Since this holds for any $z' \in R(z)$, we obtain $R(z) \subseteq R(x)$. An analogous argument shows that $R(z) \subseteq R(y)$. $\Box$

Let $\mathcal{T}_R = \mathcal{T}(\mathcal{B}_R)$. Intuitively, for each $x \in X$, the worlds accessible from x form a neighbourhood of x. In fact, they form the smallest neighbourhood of x, in the following sense:

Exercise 17

Every open neighbourhood of x in $\mathcal{T}_R$ contains $R(x)$.

The way of transforming a reflexive and transitive frame into a topological space turns out to preserve the truth values of modal formulas.

Proposition 22. *Let $M = (X, R, v)$ be a reflexive and transitive (relational) model and let $M' = (X, \mathcal{T}_R, v)$ be the corresponding topological model. Then for every formula φ in the basic modal language, and each point $x \in X$, we have $(M, x) \models \varphi$ iff $(M', x) \models \varphi$.*

Proof. Of course, the proof proceeds by structural induction. The base case where $\varphi \in$ PROP and the inductive steps for negation and conjunction are trivial.

So suppose inductively that the result holds for φ. We wish to show it holds for $\Box\varphi$. We know that $(M, x) \models \Box\varphi$ iff $R(x) \subseteq [\![\varphi]\!]_M$. By the inductive hypothesis, $[\![\varphi]\!]_M = [\![\varphi]\!]_{M'}$. Now $R(x) \subseteq [\![\varphi]\!]_{M'}$ clearly implies that $x \in int([\![\varphi]\!]_{M'})$. Conversely, if $x \in int([\![\varphi]\!]_{M'})$, then there is an open set $U \in \mathcal{T}_R$ such that $x \in U \subseteq [\![\varphi]\!]_{M'}$; by Exercise 17, we must have $R(x) \subseteq U$, which yields $R(x) \subseteq [\![\varphi]\!]_{M'}$. Thus we have shown that $(M, x) \models \Box\varphi$ iff $x \in int([\![\varphi]\!]_{M'})$; since this is in turn equivalent to $(M', x) \models \Box\varphi$, we are done. $\square$

Corollary 23. S4 *is complete with respect to the class of all topological spaces.*

Proof. Suppose $\not\vdash_{\mathsf{S4}} \varphi$. Then there is some reflexive and transitive model M that refutes φ. Then by Proposition 22 the corresponding topological model M' must also refute φ. $\square$

So S4 is sound and complete with respect to the class of all topological spaces.[111] Although we have already proved this result, it is fun and illuminating to try to prove completeness "directly" using a canonical model construction. In fact, almost all of the canonical model construction for the relational case can be salvaged. The only real difference is that instead of defining $\tilde{R}$, we must define $\tilde{\mathcal{T}}$, the canonical topology (on the same set of worlds constructed from maximally S4-consistent sets).

[111] It turns out—though this is harder to prove—that S4 is also sound and complete with respect to much more specialized classes of spaces, like $\{\mathbb{R}\}$!

Try to play around on your own with how this might work. Which sets should be open? For some guiding intuition, think about the Truth Lemma (which we'll still want to prove), and the fact that in a general topological space, open sets are just fixed points of the interior operator. And remember that to define a topology you only really need to define a basis for the topology. Have fun! (I'm happy to help you through this construction/proof if you're interested.)