Notes for formal methods seminar

Part I. Logic

Richard Pettigrew
Department of Philosophy
University of Bristol

Chapter 1

Basic set theory

Those familiar with basic set-theoretical concepts and notation should skip this chapter. Even those who are not might profitably skip it and return here when they encounter notation they don't understand.

1.1 Notation

A set and its members

- A *set* is a collection of objects.
- The objects in a set are called its *elements* or *members*.
- If S is a set and a is a member of S, then we write $a \in S$.
- If S is a set and a is *not* a member of S, then we write $a \notin S$.

Some sets of particular interest

- We write $\varnothing$ to denote the *empty set*. Thus, for all a, $a \notin \varnothing$.
- If S and T are both sets, we say that T is a subset of S if every element of T is an element of S: in this case, we write $T \subseteq S$. Note that, for every set S, $\varnothing \subseteq S$ and $S \subseteq S$.
- If a_1, a_2, ..., a_n are objects, we write $\{a_1, a_2 ..., a_n\}$ for the set whose members are all and only the a_is.
- Suppose S is a set and Φ is a property. Then we write $\{x \in S : \Phi(x)\}$ to denote the subset of S whose elements are all and only the elements of S that have the property Φ. Thus, for all a,

  $$a \in \{x \in S : \Phi(x)\} \text{ iff } (a \in S \text{ and } \Phi(a))$$

- If S is a set, we write $\wp(S)$ to denote its *power set*: that is, the set that contains all and only the subsets of S. Thus,

  $$T \in \wp(S) \text{ iff } T \subseteq S$$

  In particular, for any property Φ,

  $$\{x \in S : \Phi(x)\} \in \wp(S).$$

  But, for an infinite set S, there are many members of $\wp(S)$ that are not of the form $\{x \in S : \Phi(x)\}$.

Ordered n-tuples and Cartesian products

- If a and b are objects, we write (a, b) to denote the *ordered pair* whose first element is a and whose second element is b.
- Extending this notation, (a, b, c) is the *ordered triple* whose first element is a, second is b, and third is c.
- Generalizing this notation, $(a_1, ..., a_n)$ is the *ordered n-tuple* whose first element is a_1, whose second is a_2, whose third is a_3, and so on.
- If S and T are sets, then $S \times T$ is the set of all ordered pairs (a, b) such that $a \in S$ and $b \in T$.
- If S_1, ..., S_n are sets, then $S_1 \times ... \times S_n$ is the set of all ordered n-tuples $(a_1, ..., a_n)$ such that $a_1 \in S_1$, ..., $a_n \in S_n$. We call $S_1 \times ... \times S_n$ the *Cartesian product of* S_1, ..., S_n.
- If S is a set, then $S^n = S \times ... \times S$. That is, S^n is the set of ordered n-tuples $(a_1, ..., a_n)$ such that $a_1, ..., a_n \in S$.

Relations

- An n-place relation over a set S is a subset of S^n: that is, it is a set of n-tuples whose elements lie in S.
- We have names for various types of two-place (or *binary*) relation. Suppose R is a two-place relation over S:
  - R is *reflexive* iff $(\forall x \in S)Rxx$
  - R is *symmetric* iff $(\forall x, y \in S)(Rxy \supset Ryx)$.
  - R is *transitive* iff $(\forall x, y, z \in S)((Rxy \; \& \; Ryz) \supset Rxz)$

- We say that a binary relation R is an *equivalence relation* if R is reflexive, symmetric, and transitive.

- Given an equivalence relation R over a set S, and given $a \in S$, we define the *equivalence class of* a as follows:

$$[a] = \{x \in S : Rax\}$$

- It is a theorem that, if R is an equivalence relation over S, then the equivalence classes of R *partitions* S into mutually exclusive subsets. That is:

  - for all $a \in S$, there is $b \in S$ (namely, a itself) such that $a \in [b]$; *and*
  - for all $a, b \in S$ either $[a] = [b]$ or $[a] \cap [b] = \varnothing$.

Functions

- If S and T are sets, then $f : S \to T$ means that f is a function from S into T. That is, for each $a \in S$, $f(a) \in T$. We call $f(a)$ the *value of* f *at* a.

- If $f : S \to T$, then S is called the *domain of* f and T is called the *codomain of* f.

- If $f : S \to T$, then the *range of* f is the subset of T that contains the values of f at the various elements of S. That is, the range of f is

$$\{x \in T : \text{there is } y \in S \text{ such that } f(y) = x\}$$

- If $f : S \to T$, then we say that f *is one-one* or *injective* if, for every $a, b \in S$, if $a \neq b$, then $f(a) \neq f(b)$.

- If $f : S \to T$, then we say that f *is onto* or *surjective* if, for every $b \in T$, there is $a \in S$ such that $f(a) = b$. That is, the range of f is T.

- If $f : S \to T$, then we say that f *is a one-one correspondence* or *bijective* if f is one-one and onto: that is, injective and surjective.

Chapter 2

What is a logic?

In this part of the notes, we will discuss four logics: classical propositional logic, classical first-order logic, modal propositional logic, and classical second-order logic. In chapters 3, 4, 6, and **??**, we will present them. In the case of the former three, we will prove two central results about each, namely, their soundness and completeness theorems. In chapter 5, we will present some further results about classical first-order logic that reveal its weaknesses. In chapter **??**, we will see how these may be overcome, but at a serious cost. This will build on work on the logical study of arithmetic presented in chapter **??**.

To begin, we give a very brief definition of what will count as a logic. At first, it will seem very general and rather abstract. However, when we see the definition of the various logics in chapters 3, 4, 6, and **??**, we will begin to see why the generality is desirable.

A *logic* comprises three components:

- A *language*
- A *semantics*
- A *proof theory*

The *language* of a logic consists of a set of symbols—its *alphabet*—and a set of some of the finite strings of those symbols—its *well-formed formulae*. Usually, we specify which strings are to count as well-formed formulae by specifying a *grammar* for our language; this will be the case in the systems defined below.

The *semantics* of a logic comprises three components: a specification of the sort of mathematical object that will count as a *model* relative to which the well-formed formulae of the logic can be assigned truth values; a set of *truth-values*; and a function that takes any model and any well-formed formula to a *truth-value*—this is usually called a *valuation function.* The latter allows us to define the relation $\mathcal{M} \models \varphi$ that holds between certain models and certain well-formed formulae.

The *proof theory* of a logic identifies the sets of well-formed formulae S and the well-formed formulae φ for which φ can be proved or derived from S—in symbols, $S \vdash \varphi$. There are various ways to specify this. One condition is that the way of identifying it must be 'effective'—that is, it must consist of a mechanical procedure that a computer may carry out.

Chapter 3

Classical Propositional Logic

Enough abstract definitions. Let's meet our first example of a formal system. It's the system of *classical propositional logic* (abbreviated *CPL*). Remember that, in order to specify it, we must specify its language, semantics, and proof theory.

3.1 The language

3.1.1 The alphabet

The *alphabet* of propositional logic consists of the following symbols:

1. Propositional letters p, q, r, ... (and with subscripts)

2. Propositional connectives: $\rightarrow$ and $\neg$

3. Punctuation: (and)

The propositional connectives and punctuation are the *logical symbols* of the language. The propositional letters are its *non-logical symbols*.

The *strings* of this formal system are those (finite) sequences composed of symbols from the alphabet.

Those who have studied logic before may be concerned that our language lacks the propositional connectives $\wedge$, $\vee$, and $\leftrightarrow$. However, this is not a problem since there is a natural sense in which these can be defined in terms of the symbols we do have. We do this in the next section, once we have specified the grammar for propositional logic.

3.1.2 The grammar

Next, we must say which of these strings are to count as the *well-formed formulae* of propositional logic. To do this, we describe the *grammar* of the language, which tells us how symbols may be combined in a well-formed way. By analogy, the grammar of English tells us how the symbols of its alphabet ('a', 'b', 'c', etc.) may be combined to produce a sentence that is meaningful: for instance, it tells us that an adjective must precede a noun, and an adverb must precede or come after a verb. The grammar of propositional logic consists of similar rules.

We start by stating so-called 'closure rules' for well-formed formulae:

(1) Any propositional letter is a *well-formed formulae.*

(2) If φ and ψ are *well-formed formulae*, then $\neg\varphi$ and $(\varphi \to \psi)$ are *well-formed formulae.*

Henceforth, we will abbreviate 'well-formed formula' to 'wff'.

This tells us about a large set of strings that they are wffs: for instance, it tells us that $((p \to q) \to q)$ is a wff. (Why?) But it doesn't tell us of *any* string that it is not a wff: for instance, it doesn't tell us that $p \to)$ is not a wff. To stipulate this, we want to say that all formulae are obtained by starting with the propositional letters in (1) and applying the rules of generation in (2). We do this by saying the following: the set of wffs is the smallest set of strings of which (1) and (2) are true.

An important point: We define the set of wffs here by saying that it is the smallest set that includes certain things (the propositional letters) is closed under certain operations (the operations of negation and conditional) and is the smallest set with those properties. This mode of definition is known as *recursive definition*. It is sufficient to define the natural numbers: the set of natural numbers is the smallest set that includes zero and is closed under the operation of adding one. Thus, the theory of the language of propositional logic is as strong as the theory of arithmetic. This is important, since logical facts are often cited in attempts to provide a foundation for arithmetic.

Another consequence of this definition of the set of well-formed formulae is that we can prove theorems about them using mathematical induction. Suppose I can prove the following two facts:

(a) Each propositional letter has property Φ

(b) If wffs φ and ψ have property Φ, then wffs $\neg\varphi$ and $\varphi \to \psi$ have property Φ.

Then it follows by mathematical induction that *all* wffs have property Φ.

We are now in a position to define the other familiar propositional connectives, $\wedge$, $\vee$, and $\leftrightarrow$. By definition, if φ and ψ are wffs, we let

$$\begin{array}{rcl} (\varphi \wedge \psi) & =_{df.} & \neg(\varphi \to \neg\psi) \\ (\varphi \vee \psi) & =_{df.} & (\neg\varphi \to \psi) \\ (\varphi \leftrightarrow \psi) & =_{df.} & ((\varphi \to \psi) \wedge (\psi \to \varphi)) \end{array}$$

When we come to describe the semantics of propositional logic, we will see that the wffs on the right-hand side of these equations have the requisite semantic properties.

3.2 The semantics

In the previous section, we described the language of classical propositional logic (abbreviated 'CPL'). In this section, we describe its semantics. As noted above, we need to say what sort of mathematical object counts as a *model*, we need to specify a truth-value set, and we need to define a function that takes a model and a wff to an element of the truth-value set.

In propositional logic, our truth values are True and False. Thus, our set of truth values is $\{T, F\}$.

And a model is a truth value assignment:

Definition 3.2.1 (Propositional model) *A* propositional model $\mathcal{M} = (I)$ *is a function I that takes each propositional letter to a truth value. That is,* $I : \{p, q, r, ...\} \to \{T, F\}$.

Given a propositional model $\mathcal{M} = (I)$, we extend I to produce an *interpretation*—also denoted I—which takes each wff to its truth value in the model $\mathcal{M}$.

Definition 3.2.2 (Interpretation) *Given a propositional model* $\mathcal{M} = (I)$, *we extend I to give an* interpretation *recursively as follows:*

Suppose φ and ψ are wffs and suppose further that we have already defined $I(\varphi)$ and $I(\psi)$. Then

- *Let*

$$I(\varphi \to \psi) = \begin{cases} T & \textit{if } I(\varphi) = F \textit{ or } I(\psi) = T \\ F & \textit{otherwise} \end{cases}$$

- *Let*

$$I(\neg\varphi) = \begin{cases} T & \textit{if } I(\varphi) = F \\ F & \textit{otherwise} \end{cases}$$

Some notation

(1) If φ is a wff and $\mathcal{M}$ is a propositional model, we write $\mathcal{M} \models \varphi$, if $I(\varphi) = T$. In this case, we say that φ *is true in the model* $\mathcal{M}$.

(2) We write $S \models \varphi$, if $\mathcal{M} \models \varphi$, for any model $\mathcal{M}$ such that $\mathcal{M} \models \psi$ for all ψ in S.

In this case, we say that φ *is a logical consequence of* S.

(3) We write $\models \varphi$, if φ is a wff and $\mathcal{M} \models \varphi$ for any propositional model $\mathcal{M}$.

In this case, we say that φ *is logically valid*.

(4) If S is a set of wffs and $\mathcal{M}$ is a model,we write $\mathcal{M} \models S$ if $\mathcal{M} \models \varphi$ for all φ in S.

(5) If S is a set of wffs, we say that S *is satisfiable* if there is a propositional model $\mathcal{M}$ such that $\mathcal{M} \models S$.

(6) We say that S *is finitely satisfiable* if every finite subset of S is satisfiable. That is, for all finite $T \subseteq S$, there is $\mathcal{M}$ such that $\mathcal{M} \models T$.

(7) Given a propositional model $\mathcal{M}$, let

$$\text{Th}(\mathcal{M}) = \{\varphi : \varphi \text{ is a wff and } \mathcal{M} \models \varphi\}$$

Th($\mathcal{M}$) is called the *theory of the model* $\mathcal{M}$.

3.3 The proof theory

Many different combinations of sets of axioms and sets of rules of inference give rise to the same proof relation $\vdash$ in propositional logic. We choose a particular set. We will present them in the course of giving a definition of a *proof* of a wff φ from a set of assumptions S.

Definition 3.3.1 (Proof) *Suppose S is a set of wffs and suppose that $P = (\varphi_1, ..., \varphi_n)$ is a sequence of wffs. Then we say that P* is a proof of φ_n from assumptions S *if, for every φ_i in P, one of the following is true:*

(1) φ_i *is in* S,

(2) φ_i *is an* axiom. *That is,* φ_i *is one of the following*

- $(\varphi \to (\psi \to \varphi))$, *for some wffs* φ *and* ψ,
- $((\varphi \to (\psi \to \theta)) \to ((\varphi \to \psi) \to (\varphi \to \theta)))$, *for some wffs* φ, ψ, *and* θ,
- $((\neg\psi \to \neg\varphi) \to (\varphi \to \psi))$, *for some wffs* φ *and* ψ, *or*

(3) *There is* $j_1, j_2 < i$ *such that* φ_{j_1} *is* $\psi \to \varphi_i$ *and* φ_{j_2} *is* ψ. *In this case, we say that* φ_i is inferred from φ_{j_1} and φ_{j_2} by *modus ponens*.

Some notation

(1) We write $S \vdash_{CPL} \varphi$ if there is a proof of φ from assumptions S. In this case, we say that φ *is a syntactic consequence of* S.

(2) We write $\vdash_{CPL} \varphi$ if there is a proof of φ from the empty set of assumptions, $\varnothing$. In this case, we say that φ is a *theorem of propositional logic.*

(3) We say that S *is consistent* if is there is no proof of $\bot =_{df.} \neg(p \to (p \to p))$ from assumptions S; we say that S *is inconsistent* if there is such a proof.

Throughout this chapter, we will drop the subscript on $\vdash_{CPL}$, since we will always be talking about proofs within classical propositional logic. However, in future chapters, when we wish to compare provability in different systems, we might introduce it again.

We complete our discussion of the syntax of our system by stating the following important theorems about our formal system:

Theorem 3.3.2 *If* S *is inconsistent, then* $S \vdash \varphi$, *for any wff* φ.

Proof. Suppose $S \vdash \bot =_{df.} \neg(p \to (p \to p))$. Then

S	$\vdash$	$\neg(p \to (p \to p))$	Assumption
S	$\vdash$	$\neg(p \to (p \to p)) \to (\neg\varphi \to \neg(p \to (p \to p)))$	Axiom
S	$\vdash$	$(\neg\varphi \to \neg(p \to (p \to p)))$	Modus ponens
S	$\vdash$	$(\neg\varphi \to \neg(p \to (p \to p))) \to ((p \to (p \to p)) \to \varphi)$	Axiom
S	$\vdash$	$((p \to (p \to p)) \to \varphi)$	Modus ponens
S	$\vdash$	$(p \to (p \to p))$	Axiom
S	$\vdash$	φ	Modus ponens

Theorem 3.3.3 (Deduction Theorem) *If* $S \cup \{\varphi\} \vdash \psi$*, then* $S \vdash \varphi \rightarrow \psi$*.*

Theorem 3.3.4 (Proof by Contradiction) *If* $S \cup \{\neg\varphi\} \vdash \bot$*, then* $S \vdash \varphi$*.*

Proof. By the Deduction Theorem, if $S \cup \{\neg\varphi\} \vdash \bot$, then $S \vdash \neg\varphi \rightarrow \bot$.

S	$\vdash$	$\neg\varphi \rightarrow \neg(p \rightarrow (p \rightarrow p))$	Assumption
S	$\vdash$	$(\neg\varphi \rightarrow \neg(p \rightarrow (p \rightarrow p))) \rightarrow ((p \rightarrow (p \rightarrow p)) \rightarrow \varphi)$	Axiom
S	$\vdash$	$((p \rightarrow (p \rightarrow p)) \rightarrow \varphi)$	Modus ponens
S	$\vdash$	$(p \rightarrow (p \rightarrow p))$	Axiom
S	$\vdash$	φ	Modus ponens

as required. □

3.4 Questions

(i) Show that $(p \rightarrow (q \rightarrow p))$ is a wff of propositional logic.

(ii) Show, using mathematical induction, that no wff of propositional logic includes $\rightarrow)$ within it.

(iii) Show that $(p \rightarrow)$ is not a wff.

(iv) Show, by appealing to mathematical induction, that $(p \rightarrow qp)$ is not a wff.

(v) Show that

$$I(\varphi \wedge \psi) = \begin{cases} T & \text{if } I(\varphi) = T \text{ and } I(\psi) = T \\ F & \text{otherwise} \end{cases}$$

where $\varphi \wedge \psi$ is as defined above. Do likewise for $\varphi \vee \psi$ and $\varphi \leftrightarrow \psi$.

3.5 The soundness and completeness theorems

This completes our description of the language, semantics, and proof theory of propositional logic. We turn now to the metatheory. There are three sorts of results in the metatheory of a formal system: those that concern the proof theory of the system, those that concern its model theory, and those that establish the links between the two. We begin with the two fundamental results in the latter category: the Soundness and Completeness Theorems.

Definition 3.5.1 (Soundness) *We say that a logic is* sound *with respect to a semantics if all syntactic consequences are also semantic consequences. That is, if*

$$S \vdash \varphi \qquad \textit{implies} \qquad S \models \varphi$$

Definition 3.5.2 (Completeness) *We say that a logic is* complete *with respect to a semantics if every semantic consequence is a syntactic consequence. That is, if*

$$S \models \varphi \qquad \textit{implies} \qquad S \vdash \varphi$$

It is the task of the next two sections to prove that propositional logic is sound and complete with respect to the semantics we have just described: that is, a wff φ can be derived from a set of assumptions S if, and only if, φ is true in all models in which each wff in S is true.

3.5.1 Soundness

Theorem 3.5.3 (Soundness theorem for propositional logic) *If S is a set of wffs and φ is a single wff, then*

$$S \vdash \varphi \qquad \textit{implies} \qquad S \models \varphi$$

Proof. We will prove the following:

> For any n, if $(\varphi_1, ..., \varphi_n)$ is a proof of φ_n from a set of assumptions S, then $S \models \varphi$.

We call this our *inductive hypothesis*.

Our proof then has two stages:

- First, we will show that the inductive hypothesis holds for $n = 1$: this is called the BASE CASE.
- Then, we will show that, if it holds for all $k < n$, then it holds for n: this is called the INDUCTIVE STEP.

Proofs that employ this strategy are called *proofs by mathematical induction.*

BASE CASE We show that, if $P = (\varphi_1)$ is a proof of φ_1 from assumptions S, then $S \models \varphi_1$. If $P = (\varphi_1)$ is a proof of φ_1 from assumptions S, then either φ_1 is an axiom or φ_1 is in S.

- Suppose φ_1 is an axiom. It is easy to see that $\mathcal{M} \models \varphi_1$, for any model $\mathcal{M}$. (Try it!) So *a fortiori* $S \models \varphi$, as required.
- Suppose φ_1 is in S. Then clearly, if $\mathcal{M} \models S$, then $\mathcal{M} \models \varphi_1$. That is, $S \models \varphi_1$, as required.

INDUCTIVE STEP Assume the following: For any $k < n$ and any set of wffs S', if $P' = (\varphi'_1, ..., \varphi'_k)$ is a proof of φ'_k from assumptions S', then $S' \models \varphi'_k$. We call this the *inductive hypothesis*. We now prove the following: For any set of wffs S, if $P = (\varphi_1, ..., \varphi_n)$ is a proof of φ_n from assumptions S, then $S \models \varphi_n$.

This time there are three options: φ_n is an axiom, φ_n is in S, or φ_n is derived from φ_i and φ_j by *modus ponens*. The first two options are dealt with exactly as in the BASE CASE. Thus, we turn to the third options:

- Suppose $i, j < n$ and φ_n is derived from φ_i and φ_j by *modus ponens*: that is, $\varphi_i = \psi \to \varphi_n$ and $\varphi_j = \psi$. Then, $(\varphi_1, ..., \varphi_i)$ is a proof of φ_i from assumptions S and $(\varphi_1, ..., \varphi_j)$ is a proof of φ_j from assumptions S. Thus, by the inductive hypothesis, $S \models \psi \to \varphi_n$ and $S \models \psi$. It follows immediately that $S \models \varphi_n$ as required.

This completes our induction. □

3.5.2 Completeness

Theorem 3.5.4 (Completeness theorem for propositional logic) *If S is a set of wffs and φ is a single wff, then*

$$S \models \varphi \quad \textit{implies} \quad S \vdash \varphi$$

Our strategy is not to prove the theorem directly. Rather, we're going to state a theorem that is equivalent to it (Theorem 3.5.5), prove that it is equivalent (Proposition 3.5.6), and prove that instead. Here is the theorem that is equivalent to the completeness theorem.

Theorem 3.5.5 (The alternative Completeness theorem) *Suppose S is a set of wffs. Then if S is consistent, then S is satisfiable.*

And here is the proof that it is equivalent to the completeness theorem.

Proposition 3.5.6 *Theorems 3.5.4 and 3.5.5 are equivalent.*

Proof. First, we assume that Theorem 3.5.4 is true and prove that Theorem 3.5.5 follows. Then, we assume that Theorem 3.5.5 is true and prove that Theorem 3.5.4 follows.

- Suppose Theorem 3.5.4 is true. We want to show that Theorem 3.5.5 follows. To that end, suppose that S is consistent. We must show that there is a model $\mathcal{M}$ such that $\mathcal{M} \models S$.

  S is consistent. Thus, $S \nvdash \bot$. Thus, by the contrapositive of Theorem 3.5.4, it follows that $S \not\models \bot$. That is, it is not the case that every model that makes S true also makes $\bot$ true. Thus, there is a model in which S is true and $\bot$ is false. *A fortiori*, there is a model in which S is true, as required.

  Thus, Theorem 3.5.4 entails Theorem 3.5.5.

- Now, suppose Theorem 3.5.5 holds. And suppose that $S \models \varphi$. Then there is no model of $S \cup \{\neg\varphi\}$. Thus, by the contrapositive to Theorem 3.5.5, $S \cup \{\neg\varphi\}$ is not consistent. That is,

  $$S \cup \{\neg\varphi\} \vdash \bot$$

  It follows from this, by Theorem 3.3.4 above, that

  $$S \vdash \varphi$$

  as required. Thus, Theorem 3.5.5 entails Theorem 3.5.4.

Thus, Theorem 3.5.4 and Theorem 3.5.5 are equivalent. □

It follows that, in order to prove Theorem 3.5.4, it will suffice to prove Theorem 3.5.5.

The structure of the proof

In this section, we describe the structure of the proof of Theorem 3.5.5. In the next section, we fill in the details.

The proof proceeds by proving the following two propositions:

Proposition 3.5.7 (Lindenbaum Lemma for Propositional Logic) *If S is a consistent set of wffs, then there is a set S_∞ of wffs with the following properties:*

(a) $S \subseteq S_\infty$.

  (*That is, every formula in S is in S_∞.*)

(b) S_∞ *is maximal.*

  (*That is, for every wff ψ, either ψ is in S_∞ or $\neg\psi$ is in S_∞.*)

(c) S_∞ *is consistent.*
(*That is,* $S_\infty \nvdash \bot$)

Proposition 3.5.8 *If S_∞ is a maximal consistent set of wffs, then there is a propositional model $\mathcal{M}$ such that, for any wff ψ,*

$$\mathcal{M} \models \psi \quad \textit{iff} \quad \psi \textit{ is in } S_\infty$$

We might consider Proposition 3.5.7 as laying the foundations on which the model will be built. In Proposition 3.5.8, we build the model on top of those foundations.

Laying the foundations

Proof of Proposition 3.5.7. Suppose S is a consistent set of wffs. We are going to construct S_∞ by augmenting S infinitely many times, as follows.

1. First, we list *all* the wffs of propositional logic: φ_1, φ_2, φ_3, ...
2. Then, we define S_n recursively as follows:
   - First, we let
   $$S_0 = S$$
   - Second, we let
   $$S_{n+1} = \begin{cases} S_n \cup \{\varphi_n\} & \text{if } S_n \cup \{\varphi_n\} \text{ is consistent} \\ S_n \cup \{\neg\varphi_n\} & \text{otherwise} \end{cases}$$
3. Finally, we let
$$S_\infty = \bigcup_{n=0}^{\infty} S_n = \{\varphi : \varphi \text{ is in } S_n \text{ for some } n\}$$

To complete our proof, we need to show that S_∞, which we just constructed, is maximal, consistent, and contains S. By construction, it contains S and is maximal. Thus, we need only show that S_∞ is consistent. Suppose not. Then $S_\infty \vdash \bot$. But since proofs are finite sequences of formulae, there must be a finite subset S'_∞ of S_∞ such that $S'_\infty \vdash \bot$. But then there is n such that $S_n \vdash \bot$. But it is easy to prove by induction that, if S is consistent, then each S_n is consistent: by assumption, $S_0 = S$ is consistent; and, if S_n is consistent, then either $S \cup \{\varphi_n\}$ or $S \cup \{\neg\varphi_n\}$ is consistent (or possibly both). Thus, we have a contradiction. So, S_∞ is consistent. This completes the proof of Proposition 3.5.7. We have laid the foundations for our model.

□

Building the model

We must turn now to proving Proposition 3.5.8; that is, we must now build the model on the foundations we have laid.

Suppose S_∞ is a maximal consistent set of wffs. We now use S_∞ to construct a propositional model $\mathcal{M} = (I)$ with the following property: for all wffs ψ,

$$\mathcal{M} \models \psi \text{ iff } \psi \text{ is in } S_\infty$$

We define I as follows: for all propositional letters p,

$$I(p) = \begin{cases} T & \text{if } p \text{ is in } S_\infty \\ F & \text{otherwise} \end{cases}$$

We now prove that, when we extend I recursively to give an interpretation—also denoted I—we get that, for all wffs ψ,

$$I(\psi) = T \text{ iff } \psi \text{ is in } S_\infty$$

We proceed by induction on the complexity of ψ:

Base Case If ψ is a propositional letter, then, by definition, $I(\psi) = T$ iff ψ is in S_∞.

Inductive Step There are two cases:

- ψ is $\neg\varphi$.
  - (i) First, we show that, if $I(\neg\varphi) = T$, then $\neg\varphi$ is in S_∞. Suppose $I(\neg\varphi) = T$. Then $I(\varphi) = F$. By inductive hypothesis, it follows that φ is not in S_∞. By maximality of S_∞, it follows that $\neg\varphi$ is in S_∞, as required.
  - (ii) Second, we show that, if $\neg\varphi$ is in S_∞, then $I(\neg\varphi) = T$. Suppose $\neg\varphi$ is in S_∞. Then, by consistency of S_∞, φ is not in S_∞, since

    $$\varphi, \neg\varphi \vdash \bot$$

    for any wff φ. (Show this!) Thus, by inductive hypothesis, $I(\varphi) = F$. So $I(\neg\varphi) = T$.
- ψ is $(\varphi \to \theta)$.
  - (i) First, we show that, if $I(\varphi \to \theta) = T$, then $(\varphi \to \theta)$ is in S_∞. Suppose $I(\varphi \to \theta) = T$. Then $I(\varphi) = F$ or $I(\theta) = T$. By inductive hypothesis, it follows that φ is not in S_∞ or θ is in

S_∞. Suppose $\neg(\varphi \to \theta)$ is in S_∞. Then we have that S_∞ is inconsistent, since

$$\neg\varphi, \neg(\varphi \to \theta) \vdash \bot \quad \text{and} \quad \theta, \neg(\varphi \to \theta) \vdash \bot$$

This contradicts the consistency of S_∞. Thus, by maximality of S_∞, $(\varphi \to \theta)$ is in S_∞.

(ii) Second, we show that if $\varphi \to \theta$ is in S_∞, then $I(\varphi \to \theta) = T$. Suppose $\varphi \to \theta$ is in S_∞. Then, since S_∞ is consistent, it cannot be the case that φ and $\neg\theta$ are in S_∞ since

$$\varphi, \neg\theta, \varphi \to \theta \vdash \bot$$

Thus, φ is not in S_∞ or $\neg\theta$ is not in S_∞. By inductive hypothesis, it follows that $I(\varphi) = F$ or $I(\theta) = T$. In either case, $I(\varphi \to \theta) = T$, as required.

This completes our proof of Proposition 3.5.8, and that completes our proof of the Completeness Theorem for Propositional Logic.

Chapter 4

Classical first-order logic

In propositional logic, we represent propositions as truth-functional combinations of further propositions. In first-order logic, we delve a little more deeply into the logical structure of the propositional letters out of which all propositional logic are ultimately constructed. As before, we meet the language first, then the semantics and the proof theory in that order. Finally we marry the two with soundness and completeness theorems.

4.1 The language

4.1.1 The alphabet

The *alphabet* of first-order logic consists of the following symbols:

1. Constants: **a**, **b**, **c**, ... (and with subscripts)
2. Free variables: a, b, c, ... (and with subscripts)
3. Bound variables: u, v, w, x, y, z (and with subscripts)
4. Function symbols: f^1, f^2, ..., g^1, g^2, ... (and with subscripts)
5. Relation symbols: P^0, P^1, P^2, ..., R^0, R^1, R^2, ... (and with subscripts)
6. Propositional connectives: $\rightarrow$ and $\neg$
7. The universal quantifier: $\forall$
8. Punctuation: (and) and ,

The superscripts on the function and relation symbols indicate their *arity*. Thus, f^n is an n-ary function symbol. That is, it represents an n-place function. Similarly, P^n is an n-ary relation symbol. That is, it represents an n-place relation.

The quantifier, propositional connectives, punctuation, and free and bound variables are the *logical symbols* of the language. The constants, function, and relation symbols are its *non-logical symbols*.

Definition 4.1.1 (Signature) *A first-order* signature *is a set of non-logical symbols: that is, a set of constant symbols, function symbols, and relation symbols.*

The *strings* of this formal system are those (finite) sequences composed of symbols from the alphabet. Given a signature σ, the σ*-strings* are the strings whose non-logical symbols are drawn from σ.

4.1.2 The grammar

Next, we must say which of these strings are to count as the *well-formed formulae* of first-order logic. These are considerably more complicated than in the propositional case.

Terms First, we identify a set of strings that we call the *terms* of the language. Roughly, these are the naming phrases of the language; they are the strings that are used to denote objects.

Definition 4.1.2 (Term)

(1) *Every constant and every free variable is a* term*.*

(2) *If* $t_1, ..., t_n$ *are* terms *and* f^n *is a function symbol, then* $f^n(t_1, ..., t_n)$ *is a* term*.*

(3) *There are no other* terms*.*

Note that, in (2), the number of terms, t_1, ..., t_n must match the superscript on the function symbol. As noted above, we call that superscript the *arity* of the function symbol.

Atomic formulae Next, we pick out the *atomic formulae* of the language. These will be the building blocks of the well-formed formulae. Roughly, they assert that a particular relation holds between particular objects.

Definition 4.1.3 (Atomic formula)

(1) *If* t_1, ..., t_n *are terms and* R^n *is a relation symbol, then* $R^n(t_1, ..., t_n)$ *is an* atomic formula.

(2) *There are no other* atomic formulae.

As in the definition of terms above (Definition 4.1.2), note that the number of terms, t_1, ..., t_n must match the superscript on the relation symbol. Again, we call that superscript the *arity* of the relation symbol.

Well-formed formulae Finally, we are in a position to pick out the *well-formed formulae* from amongst the strings of symbols of the first-order language.

Definition 4.1.4 (Well-formed formula) *The set of well-formed formulae (or wffs) of first-order logic is the smallest set that has the following closure properties:*

(1) *Atomic formulae are* well-formed formulae.

(2) *If* φ *and* ψ *are* well-formed formulae, *then* $\neg\varphi$ *and* $(\varphi \rightarrow \psi)$ *are* well-formed formulae.

(3) *If* φ *is a* well-formed formulae, *then* $\forall x \varphi[x/a]$ *is a* well-formed formula, *where* $\varphi[x/a]$ *is the string obtained from the well-formed formula* φ *by substituting the bound variable* x *in place of the free variable* a *throughout.*

Definition 4.1.5 (σ-wff) *Given a signature* σ, *a* σ-wff *is a wff whose non-logical symbols are drawn from* σ.

Those who have studied logic before may be concerned that our language lacks the existential quantifier $\exists$. However, as in the case of the propositional connectives, this is not a problem since there is a natural sense in which this can be defined in terms of the symbols we do have. Thus, by definition, if φ is a wff, we let

$$\exists x \varphi[x/a] \quad =_{df.} \quad \neg\forall x \neg\varphi[x/a]$$

When we come to describe the semantics of first-order logic, we will see that the wff on the right-hand side of this equation has the requisite semantic properties.

4.2 The semantics

In the previous section, we described the language of classical first-order logic (FOL). As noted above, we now need to say what sort of mathematical object counts as a *model* over a given signature σ, we need to specify a truth-value set, and we need to define a function that takes a σ-model and a σ-wff to an element of the truth-value set.

4.2.1 First-order models

Definition 4.2.1 (σ-model) *Given a signature σ, a* σ-model *is a pair, $\mathcal{M} = (D, I)$, where D is a set and I is a function defined as follows:*

(1) *For each constant symbol $\mathbf{c}$ in σ, $I(\mathbf{c}) \in D$.*

(2) *For each free variable a, $I(a) \in D$.*

(3) *For each function symbol f^n in σ, $I(f^n) : D^n \to D$.*

(4) *For each relation symbol R^n in σ, $I(R^n) \subseteq D^n$.*

We call D the domain of $\mathcal{M}$ *and we call I the* interpretation function of $\mathcal{M}$.

4.2.2 Truth-value set

We have said what it is to be a model of first-order logic. Before we can complete our description of first-order semantics, we must say what the truth-value set is. It is $\{T, F\}$: that is, it contains only two truth-values, True and False.

4.2.3 Extending the interpretation function

With this in hand, we can complete our description of the semantics for first-order logic by saying, given a σ-wff φ and a model $\mathcal{M}$ over signature σ, which truth-value is assigned to φ interpreted in $\mathcal{M}$.

We do this by extending the interpretation function I of $\mathcal{M}$ in three stages: by the third stage, given a wff φ, there is a value $I(\varphi) \in \{T, F\}$.

- **First step** First, we extend I so that, for each term t, it gives $I(t) \in D$.

  (i) We already have that, for any constant symbol $\mathbf{c}$ or free variable a, $I(\mathbf{c}) \in D$ and $I(a) \in D$.

(ii) Now suppose t_1, ..., t_n are terms and f^n is a function symbol. And suppose that we have already extended I to give values for $I(t_1)$, ..., $I(t_n) \in D$. Then let

$$I(f^n(t_1, ..., t_n)) =_{df.} I(f^n)(I(t_1), ..., I(t_n))$$

That is, given $f^n(t_1, ..., t_n)$, I takes the n-tuple $(I(t_1), ..., I(t_n))$ and applies the function $I(f^n)$ to it.

- **Second step** Next, we extend I so that, for each atomic formula φ, it gives $I(\varphi) \in \{T, F\}$. Suppose t_1, ..., t_n are terms and R^n is a relation symbol. Suppose further that we have already extended I to give values for $I(t_1)$, ..., $I(t_n) \in D$. Then let

$$I(R^n(t_1, ..., t_n)) =_{df.} \begin{cases} T & \text{if } (I(t_1), ..., I(t_n)) \in I(R^n) \\ F & \text{if } (I(t_1), ..., I(t_n)) \notin I(R^n) \end{cases}$$

- **Third step** Finally, we extend I so that, for each wff φ, it gives $I(\varphi) \in \{T, F\}$. Before we do this, we need to introduce some notation: Given an interpretation function I, a free variable a, and $d \in D$, let I_a^d be the interpretation function obtained from I be setting $I(a) = d$.

  Now, suppose φ and ψ are wffs and suppose further that we have already extended I to give values for $I(\varphi)$ and $I(\psi) \in \{T, F\}$. Then

  (i) Let

$$I(\varphi \to \psi) =_{df.} \begin{cases} T & \text{if } I(\varphi) = F \text{ or } I(\psi) = T \\ F & \text{otherwise} \end{cases}$$

  (ii) Let

$$I(\neg\varphi) =_{df.} \begin{cases} T & \text{if } I(\varphi) = F \\ F & \text{otherwise} \end{cases}$$

  (iii) Let

$$I(\forall x \varphi[x/a]) =_{df.} \begin{cases} T & \text{if, for every } d \in D, I_a^d(\varphi) = T \\ F & \text{otherwise} \end{cases}$$

Definition 4.2.2 *Suppose $\mathcal{M} = (D, I)$ is a model, φ is a wff containing free variables $a_1, ..., a_n$, and $d_1, ..., d_n$ are elements of D. Then we write*

$$\mathcal{M} \models \varphi[d_1/a_1, ..., d_n/a_n]$$

to mean that

$$\mathcal{M}_{a_1,...,a_n}^{d_1,...,d_n} \models \varphi$$

where $\mathcal{M} = (D, I_{a_1,...,a_n}^{d_1,...,d_n})$: recall that $I_{a_1,...,a_n}^{d_1,...,d_n}$ differs from I only in that it takes $I(a_i) = d_i$ for $i = 1, 2, ..., n$.

4.3 The proof theory

Many different combinations of sets of axioms and sets of rules of inference give rise to the same theorems in first-order logic. We choose a particular set. We will present them in the course of giving a definition of a *proof* of a wff φ from a set of assumptions S.

Definition 4.3.1 (Proof) *Suppose S is a set of wffs and suppose that $P = (\varphi_1, ..., \varphi_n)$ is a sequence of wffs. Then we say that P* is a proof of φ_n from assumptions S *if, for every φ_i in P, one of the following is true:*

(1) *φ_i is in S,*

(2) *φ_i is an* axiom. *That is, φ_i is one of the following*

- *$(\varphi \to (\psi \to \varphi))$, for some wffs φ and ψ,*
- *$((\varphi \to (\psi \to \theta)) \to ((\varphi \to \psi) \to (\varphi \to \theta)))$, for some wffs φ, ψ, and θ,*
- *$((\neg\psi \to \neg\varphi) \to (\varphi \to \psi))$, for some wffs φ and ψ, or*
- *$(\forall x\varphi[x/a] \to \varphi[t/a])$, for some wff φ and term t.*

(3) *There is $j_1, j_2 < i$ such that φ_{j_1} is $\psi \to \varphi_i$ and φ_{j_2} is ψ. In this case, we say that φ_i* is inferred from φ_{j_1} and φ_{j_2} by *modus ponens*.

(4) *There is $j < i$ such that φ_i is $\forall x\varphi_j[x/a]$, $(\varphi_1, ..., \varphi_j)$ is a proof of φ_j from assumptions S, and a does not occur in any wff in S. In this case, we say that φ_i* is inferred from φ_j by *Universal Generalization*.

This time, we write $S \vdash_{FOL} \varphi$ if there is a proof of φ from assumptions S in classical first-order logic. As in the propositional case, we drop the subscript throughout this chapter, but may introduce it again in future chapters.

4.4 The soundness and completeness theorems

This completes our description of the formal system of first-order logic and its semantics. We turn now to the soundness and completeness theorems.

4.4.1 Soundness

Theorem 4.4.1 (Soundness theorem for first-order logic) *If S is a set of wffs and φ is a single wff, then*

$$S \vdash \varphi \quad \textit{implies} \quad S \models \varphi$$

Proof. As in the propositional case, we will prove the following:

> For any n, if $(\varphi_1, ..., \varphi_n)$ is a proof of φ_n from a set of assumptions S, then $S \models \varphi$.

BASE CASE This is exactly as in the propositional case, except that we must show that

$$\mathcal{M} \models (\forall x \varphi[x/a] \to \varphi[t/a])$$

for any model $\mathcal{M}$, wff φ, and term t. INDUCTIVE STEP Assume the following: For any $k < n$ and any set of wffs S', if $P' = (\varphi'_1, ..., \varphi'_k)$ is a proof of φ'_k from assumptions S', then $S' \models \varphi'_k$. We call this the *inductive hypothesis.* We now prove the following: For any set of wffs S, if $P = (\varphi_1, ..., \varphi_n)$ is a proof of φ_n from assumptions S, then $S \models \varphi_n$.

This time there are four options: φ_n is an axiom, φ_n is in S, φ_n is derived from φ_i and φ_j by *modus ponens*, or φ_n is derived from φ_i by Universal Generalization. The first two options are dealt with exactly as in the BASE CASE. The third is dealt with as in the propositional case. Thus, we turn to the final option:

- Suppose $i < n$ and φ_n is inferred from φ_i by Universal Generalization. That is, φ_n is $\forall x \varphi_i[x/a]$, a does not occur in S, and $(\varphi_1, ..., \varphi_i)$ is a proof of φ_i from the assumption of S. Thus, by inductive hypothesis, $S \models \varphi_i$. So suppose $\mathcal{M} \models S$. We must show that $I(\forall x \varphi_i[x/a]) = T$. Thus, we must show that, for every $d \in D$, $I^d_a(\varphi_i) = T$. Since a does not occur in any wff in S and since $I(\psi) = T$ for every ψ in S, it follows that $I^d_a(\psi) = T$ for every ψ in S. Thus, $\mathcal{M}^d_a = (D, I^d_a) \models S$. So $\mathcal{M}^d_a = (D, I^d_a) \models \varphi_i$. Thus, $\mathcal{M} \models \forall x \varphi_i[x/a]$. That is, $\mathcal{M} \models \varphi_n$.

This completes our induction. □

4.4.2 Completeness

This result was proved first by Kurt Gödel in his doctoral thesis in 1929. The proof that we give is due to Leon Henkin who published it in 1949.

Theorem 4.4.2 (Completeness theorem for first-order logic) *If S is a set of wffs and φ is a single wff, then*

$$S \models \varphi \quad \textit{implies} \quad S \vdash \varphi$$

As in the case of propositional logic, our strategy is not to prove the theorem directly. Rather, we're going to state a theorem that is equivalent to it (Theorem 4.4.3), prove that it is equivalent, and prove that instead. Here is the theorem that is equivalent to the completeness theorem. It is the same as in the propositional case.

Theorem 4.4.3 (The alternative Completeness theorem) *Suppose S is a set of wffs of first-order logic. If S is consistent, then S is satisfiable.*

The proof that they are equivalent is exactly as in the propositional case.

The structure of the proof

The structure of the proof is exactly as in the propositional case. That is, we prove the following two lemmas:

Proposition 4.4.4 (Lindenbaum Lemma for First-order Logic) *If S is a consistent set of σ-wffs, then there is a signature σ^∞ such that $\sigma \subseteq \sigma^\infty$ and a set S_∞ of σ^∞-wfs with the following properties:*

(a) $S \subseteq S_\infty$.

(b) S_∞ *is maximal*

(c) S_∞ *is consistent.*

(d) S_∞ *is existentially witnessed.*
(*That is, if ψ is a σ^∞-wff and $\exists x\psi[x/a]$ is in S_∞, then there is a constant* $\mathbf{c}$ *in σ^∞ such that $\psi[\mathbf{c}/a]$ is in S_∞.*)

Proposition 4.4.5 *If S_∞ is a maximal, consistent, and existentially witnessed set of σ^∞-wffs, then there is a model $\mathcal{M}$ over σ^∞ such that, for any σ^∞-wff ψ,*

$$\mathcal{M} \models \psi \quad \textit{iff} \quad \psi \textit{ is in } S_\infty$$

Laying the foundations

Proof of Proposition 4.4.4. Suppose S is a consistent set of σ-wffs. We are going to construct S_∞ by extending σ to σ^∞ and then augmenting S in two steps.

1. First, we extend the signature σ by adding new constants $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$, $\cdots$, such that no $\mathbf{c}_i$ occurs in any formula in S. Call the resulting signature σ^∞.

2. Next, we list *all* the σ^∞-wffs: φ_1, φ_2, φ_3, ...

3. Then, we extend S in the following way:

   (a) $S_0 = S$

   (b) $S_{n+1} = S_n \cup \{(\exists x\varphi_{n+1}[x/a] \to \varphi_{n+1}[\mathbf{c}_j/a])\}$
   where $\mathbf{c}_j$ is the first of the new constants that does not occur in the formulae $\varphi_1, \cdots, \varphi_{n+1}$.

   Then let

   $$S^+ = \bigcup_{n=0}^{\infty} = \{\varphi : \varphi \text{ is in } S_n \text{ for some } n\}$$

   (This step guarantees that the set S_∞, which we are constructing at the moment, is existentially witnessed.)

4. Finally, beginning this time with S^+ instead of S, we construct the maximal consistent set S_∞ such that $S \subseteq S_\infty$ by the same recursive process used in the propositional case. That is, we enumerate all $\sigma^i nfty$-wffs and add either them or their negation one at a time, retaining consistency throughout.

Having constructed S_∞, we need to show that it is maximal, consistent, and existentially witnessed. The proof that it is maximal and consistent is analogous to the propositional case.

(d) Suppose $\exists x\psi[x/a]$ is in S_∞. Then, by the definition of S^+, we have that $(\exists x\psi[x/a] \to \psi[\mathbf{c}_j/a])$ is in S_∞ for some constant $\mathbf{c}_j$. Thus, $\psi[\mathbf{c}_j/a]$ is in S_∞. That is, S_∞ is existentially witnessed.

This completes the proof of Proposition 4.4.4. We have laid the foundations for our model. □

Building the model

We must turn now to proving Proposition 4.4.5; that is, we must now build the model on the foundations we have laid.

Suppose S_∞ is a maximal, consistent, and existentially witnessed set of σ^∞-wffs. We now use S_∞ to construct a model $\mathcal{M}$ over σ^∞ such that, for all σ^∞-wffs ψ

$$\mathcal{M} \models \psi \text{ iff } \psi \text{ is in } S_\infty$$

We define this model as follows: $\mathcal{M} = (D, I)$ where

- D is the set of terms that occur in S_∞.

  (Thus, each of the constants in the original signature σ will be in D; also, each new constant $\mathbf{c}_1$, $\mathbf{c}_2$, ...; also, each n-ary function symbol of σ applied to every list of n of those constants; and so on.)

- $I(\mathbf{c}) = \mathbf{c}$, if $\mathbf{c}$ is a constant symbol in σ^∞.

- $I(a) = a$, if a is a free variable.

- $I(R^n) = \{(t_1, ..., t_n) : t_1, \cdots, t_n$ are terms and $R^n(t_1, ..., t_n)$ is in $S_\infty\}$, if R^n is a relation symbol in σ^∞.

- $I(f^n)(t_1, ..., t_n) = f^n(t_1, ..., t_n)$, if f^n is a function symbol in σ^∞.

This completes our description of the model. Now we show that it has the property required of it by Proposition 4.4.5.

Showing that it does what we want

We must prove the following proposition. From this, Proposition 4.4.5 follows.

Proposition 4.4.6 *For any σ^∞-wff φ,*

$$\mathcal{M} = (D, I) \models \varphi \quad \textit{iff} \quad \varphi \textit{ is in } S_\infty$$

Proof. We prove this by induction on the construction of the formulae of S.

BASE CASE If t_1, $\cdots$, t_n are terms, and R^n is a relation symbol of arity n,

$$\begin{aligned} \mathcal{M} \models R^n(t_1, \cdots, t_n) \quad & \text{iff} \quad (t_1, \cdots, t_n) \text{ is in } I(R^n) \\ & \text{iff} \quad R^n(t_1, \cdots, t_n) \text{ is in } S_\infty \end{aligned}$$

INDUCTIVE STEP Suppose that, for all formulae φ containing n or fewer connectives and quantifiers,

$$\mathcal{M} \models \varphi \quad \text{iff} \quad \varphi \text{ is in } S_\infty$$

Call this the *inductive hypothesis.*

- Suppose $(\psi \to \theta)$ has $n+1$ connectives. Then ψ and θ have n or fewer connectives. Thus, by the inductive hypothesis,

$$\mathcal{M} \models \psi \quad \text{iff} \quad \psi \text{ is in } S_\infty$$

and

$$\mathcal{M} \models \theta \quad \text{iff} \quad \theta \text{ is in } S_\infty$$

Thus,

$$\begin{aligned} \mathcal{M} \models (\psi \to \theta) \quad & \text{iff} \quad \mathcal{M} \not\models \psi \text{ or } \mathcal{M} \models \theta \\ & \text{iff} \quad \psi \text{ is not in } S_\infty \text{ or } \theta \text{ is in } S_\infty \\ & \text{iff} \quad (\psi \to \theta) \text{ is in } S_\infty \end{aligned}$$

- Suppose $\neg\psi$ has $n+1$ connectives. Then ψ has n connectives. Thus, by the inductive hypothesis,

$$\mathcal{M} \models \psi \quad \text{iff} \quad \psi \text{ is in } S_\infty$$

Thus,

$$\begin{aligned} \mathcal{M} \models \neg\psi \quad & \text{iff} \quad \mathcal{M} \not\models \psi \\ & \text{iff} \quad \psi \text{ is not in } S_\infty \\ & \text{iff} \quad \neg\psi \text{ is in } S_\infty \end{aligned}$$

- Suppose $\forall x\psi[x/a]$ has $n+1$ connectives. Then, for any term t, $\psi[t/a]$ has n connectives. Thus, by the inductive hypothesis,

$$\mathcal{M} \models \psi[t/a] \quad \text{iff} \quad \psi[t/a] \text{ is in } S_\infty$$

Thus,

$$\begin{aligned} \mathcal{M} \models \forall x\psi[x/a] \quad & \text{iff} \quad \mathcal{M} \models \psi[t/a] \text{ for every term } t \text{ in } D \\ & \text{iff} \quad \psi[t/a] \text{ is in } S_\infty \text{ for every term } t \text{ in } D \\ & \text{iff} \quad \forall x\psi[x/a] \text{ is in } S_\infty \end{aligned}$$

The final 'iff' requires some further justification:

(i) ('If') First, suppose that $\forall x\psi[x/a]$ is in S_∞. Then, if there was a term t such that $\neg\psi[t/a]$ is also in S_∞, then S_∞ would be inconsistent. Thus, $\psi[t/a]$ is in S_∞ for every term t in D.

(ii) ('Only if') Second, suppose that $\forall x \psi[x/a]$ is not in S_∞. Then, since S_∞ is maximal, it follows that $\neg \forall x \psi[x/a]$ is in S_∞. Thus, $\exists x \neg \psi[x/a]$ is in S_∞. Since S_∞ is a maximal consistent extension of S^+, there is a constant $\mathbf{c}_j$ such that the wff $(\exists x \neg \psi[x/a] \rightarrow \neg \psi[\mathbf{c}_j/a])$ is in S_∞. It follows that $\neg \psi[\mathbf{c}_j/a]$ is in S_∞. Thus, it is not the case that $\psi[t/a]$ is in S_∞ for every t in D. By contraposition, this establishes that, if $\psi[t/a]$ is in S_∞ for every t in S, then $\forall x \psi[x/a]$ is in S_∞, as required.

This completes our proof of Proposition 4.4.6. □

This in turn completes our proof of Proposition 4.4.5. □

The alternative Completeness Theorem then follows from Proposition 4.4.5. Given a model $\mathcal{M} = (D, I)$ over σ^∞ of S_∞, we can create a σ-model $\mathcal{M}' = (D', I')$ of S: $D' = D$ and I' is the restriction of I to the signature σ. This completes our proof of the Completeness Theorem for First-Order Logic.

Chapter 5

The metatheory of first-order logic

5.0.3 Compactness

Suppose S is a set of wffs in the language of first-order logic. Recall that S is finitely satisfiable if, for every finite subset S' of S, S' is satisfiable. The compactness theorem tells us that this is sufficient to make S itself satisfiable.

Theorem 5.0.7 (Compactness theorem for first-order logic) *If S is finitely satisfiable, then S is satisfiable.*

If this doesn't surprise you on first reading it, you may wish to think a little more about it. It's an extremely surprising result; as we shall see, it has extremely surprising consequences.

Proof of Theorem 5.0.7. We proceed by proving the contrapositive of the theorem. That is, we will prove that, if S is not satisfiable, then there is a finite subset of S that is not satisfiable. Thus, suppose that S is not satisfiable. That is, there is no model $\mathcal{M}$ such that $\mathcal{M} \models S$. Thus, it follows from the alternative version of the completeness theorem (Theorem **??**) that S is inconsistent. That is, there is a proof of a contradiction from the assumption of S. Since every proof is finite, there is a finite subset S' of S such that there is a contradiction from S'. Thus, S' is inconsistent. By the soundness theorem, it follows that S' is not satisfiable. This completes the proof. □

5.0.4 Consequences of Compactness

One use to which we wish to put logic is to characterize types of structures. Let's consider two examples.

First, the following formulae characterize the structure shared by all two-element sets, in the sense that, in all models $\mathcal{M} = (D, I)$ of these formulae, D is a set containing exactly two elements:

1. $\exists x \exists y (x \neq y)$
2. $\forall x \forall y \forall z (x = y \vee y = z \vee x = z)$

Second, the following formulae characterize the *linear ordering* structures, in the sense that, in all models $\mathcal{M} = (D, I)$ of these formulae, $I(R)$ is a linear ordering of the domain D:

1. $\forall x Rxx$
2. $\forall x \forall y ((Rxy \wedge Ryx) \rightarrow x = y)$
3. $\forall x \forall y \forall z ((Rxy \wedge Ryz) \rightarrow Rxz)$
4. $\forall x \forall y (Rxy \vee Ryx)$

If we add the following formula, they characterize the *dense* linear ordering structures:

5. $\forall x \forall y (Rxy \rightarrow \exists x (Rxz \wedge Rzy))$

And if we add another formula, they characterize the dense linear orderings *with no endpoints*:

6. $\forall x (\exists y Rxy \wedge \exists z Rzx)$

The Natural Number Structure

At the end of the nineteenth century, the mathematician Richard Dedekind wished to characterize the *natural number* structure. That is, he wished to provide a set A of formulae such that $\mathcal{M} \models A$ if, and only if, $\mathcal{M}$ has the structure of the natural numbers. Since logic was still in its infancy when Dedekind wrote, he did not give his project so precise a formulation—indeed, it was another mathematician called Giuseppe Peano who, a year later, first wrote Dedekind's conditions in something that we would recognize as logical notation. However, this is certainly his intention.

However, it is a consequence of the Compactness Theorem that this aim could not be achieved using first-order logic. That is, there is no set of formulae A in first-order logic such that $\mathcal{M} \models A$ if, and only if, $\mathcal{M}$ has the structure of the natural numbers. It can be achieved using second-order logic, and that is essentially what Dedekind did in his seminal treatise *Was ist und was sollen die Zahlen?*. That is, there is a set of formulae in *second-order* logic whose models are exactly the natural number structures. But there is no such set of first-order formulae, and this is a consequence of the Compactness Theorem, as I will now show.

First, what is the natural number structure? For Dedekind, a natural number structure consisted of a set N, a distinguished zero element 0 and a successor function S, which takes one number to the next. Thus, if S is a set of wffs that hopes to characterize this structure, the signature of these wffs must include a constant symbol 0 and a one-place function symbol S. In such a signature, for any number n, we let $\overline{n}$ denote the term

$$\underbrace{SS...S}_{n}0.$$

Now suppose, for the sake of a contradiction, that the set of first-order wffs A characterizes the natural number structure. Thus, the natural numbers themselves may be used to provide a model for A: we call it $\mathbb{N} = (N, I_{\mathbb{N}})$, where N is the set of natural numbers and $I(0)$ is zero, and $I(S)$ is the function that takes a number to the number that immediately follows it. Thus, by hypothesis, $\mathbb{N} \models A$.

Next, we introduce a new constant symbol $\mathbf{c}$ and a new set of formulae, which we call A^*:

$$A^* = A \cup \{\mathbf{c} \neq \overline{n} : n = 0, 1, 2, ...\}$$

Now, there is no guarantee that $\mathbb{N} = (N, I_{\mathbb{N}}) \models A^*$. However, we can use $\mathbb{N}$ to show that every finite subset of the axioms of A^* has a model. And then we can use the Compactness Theorem to show that A^* has a model. Since any model of A^* does not have the structure of the natural numbers, and since any model of A^* is a model of A, it follows that A has a model that does not have the structure of the natural numbers.

Suppose B is a finite subset of A^*. Then let k be the greatest number such that the formula $\mathbf{c} \neq \overline{k}$ is in B. Then extend $I_{\mathbb{N}}$ to an interpretation $I^*_{\mathbb{N}}$ by letting

$$I^*_{\mathbb{N}}(\mathbf{c}) = I(\overline{k+1})$$

Then let $\mathbb{N}^* = (D, I^*_{\mathbb{N}})$. Then $\mathbb{N}^* \models B$. After all, for any $\varphi \in B$, either

- φ is in A, in which case $\mathbb{N}^* \models \varphi$, since $\mathbb{N}^* \models A$; or,
- φ is $\mathbf{c} \neq \overline{n}$ for some $n \leq k$, in which case $\mathbb{N}^* \models \varphi$, since $I^*_{\mathbb{N}}(\mathbf{c}) = I(\overline{k+1}) \neq I(\overline{n})$.

Thus, $\mathbb{N}^* \models B$, for any finite subset B of A^*. By the Compactness Theorem, it follows that A^* is satisfiable; that is, there is a model $\mathcal{M}_0 = (D_0, I_0)$ of A^*. And, thus, there is an element $I_0(\mathbf{c})$ of $\mathcal{M}_0$ such that $a \neq I_0(\overline{n})$, for all numbers n. Thus, $\mathcal{M}_0$ is *not* a natural number structure. It includes, in Dedekind's words, "alien intruders"; in particular $I_0(\mathbf{c})$. But, of course, since $\mathcal{M}_0 \models A^*$ and $A \subseteq A^*$, $\mathcal{M}_0 \models A$. Thus, it is not the case that every model of A is a natural number structure, contrary to our supposition. $\mathcal{M}_0$ is a model of A and $\mathcal{M}_0$ does not have the structure of the natural numbers. Thus, there is no first-order characterization of the natural number structure.

The Real Number Structure

A similar trick shows that there is no first-order characterization of the real number structure. Suppose that C is a set of first-order formulae that purport to give such a characterization. Amongst its signature there must be constant symbols 0 and 1, a relation symbol $<$, and function symbols $+$, $-$, $\times$, and $/$. It is a minimum requirement on C characterizing the real number structure that $\mathbb{R} = (R, I_{\mathbb{R}}) \models C$, where R is the set of real numbers, and $I_{\mathbb{R}}$ is the natural interpretation of the non-logical symbols in the signature of C.

Next, let $\overline{1/n}$ denote the term $1/\underbrace{1+1+...+1}_{n}$. Then extend the signature of C by adding a new constant $\mathbf{c}$ and let

$$C^* = C \cup \{\mathbf{c} < \overline{1/n} : n = 0, 1, 2, ...\}$$

Then, as above, suppose B is a finite subset of C^*. Let k be the greatest number such that the formula $\mathbf{c} < \overline{1/k}$ is in B. And extend $I_{\mathbb{R}}$ to $I^*_{\mathbb{R}}$ by letting

$$I^*_{\mathbb{R}}(\mathbf{c}) = I_{\mathbb{R}}(\overline{1/(k+1)})$$

then $\mathbb{R}^* \models B$. So, by Compactness, there is model $\mathcal{M}_1 = (D_1, I_1)$ of C^*. This model is also a model of C, but it does not have the structure of the real numbers since it contains an element $I_1(\mathbf{c})$, that is smaller than $I_1(\overline{1/n})$, for all $n = 1, 2, ...$ Thus, there is no first-order characterization of the real number structure.

The models $\mathcal{M}_0$ and $\mathcal{M}_1$ are known as *non-standard models* of arithmetic and real analysis respectively. The study of their structure is of great mathematical interest.

Finitude

A similar trick to that used above shows that there is no set of first-order formulae F such that $\mathcal{M} = (D, I) \models F$ if, and only if, D is a finite set. Thus, just as there is no first-order characterization of the natural or real number structures, there is no first-order characterization of finitude.

5.0.5 The Löwenheim-Skolem Theorems

In the previous section, we used the Compactness Theorem to show that there is no first-order characterization of the natural number structure, nor of the real number structure. In this section, we state the *Löwenheim-Skolem theorems*, which provide a different route to slightly weaker versions of these results.

Definition 5.0.8 (Submodels and extensions) *Suppose σ_1, σ_2 are signatures and $\sigma_1 \subseteq \sigma_2$. And suppose $\mathcal{M}_1 = (D_1, I_1)$ is a σ_1-model and $\mathcal{M}_2 = (D_2, I_2)$ is a σ_2-model. Then we say that $\mathcal{M}_1$* is a submodel of *$\mathcal{M}_2$ if*

(1) $D_1 \subseteq D_2$

(2) *For each constant c in σ_1, $I_1(c) = I_2(c)$*

(3) *For each free variable a, $I_1(a) = I_2(a)$*

(4) *For each n-place function symbol f in σ_1, $I_1(f) = I_2(f)|_{D_1^n}$, where $I_2(f)|_{D_1^n}$ is the function $I_2(f)$, whose domain is D_2^n, restricted to D_1^n.*

(5) *For each n-place relation symbol R in σ_1, $I_1(R) = I_2(R) \cap D_1^n$.*

In this case, we say that $\mathcal{M}_2$ is an extension of *$\mathcal{M}_1$.*

Definition 5.0.9 (Elementary submodels and extensions) *We say that $\mathcal{M}_1$* is an elementary submodel of *$\mathcal{M}_2$ if*

(1) *$\mathcal{M}_1$ is a submodel of $\mathcal{M}_2$*

(2) *For every σ_1-wff φ containing free variables a_1, ..., a_n and every d_1, ..., d_n in D_1*

$$\mathcal{M}_1 \models \varphi[d_1, ..., d_n] \textit{ iff } \mathcal{M}_2 \models \varphi[d_1, ..., d_n]$$

(*For the definition of this square brackets notation, see Definition 4.2.2.*)

In this case, we say that $\mathcal{M}_2$ is an elementary extension *of $\mathcal{M}_1$.*

We now state a result that gives rise to a useful test, called the *Tarski-Vaught test*, which provides necessary and sufficient conditions for two σ-models $\mathcal{M}_1$ and $\mathcal{M}_2$ to be such that $\mathcal{M}_1$ is an elementary submodel of $\mathcal{M}_2$. First, a definition.

Definition 5.0.10 *Suppose $\mathcal{M} = (D, I)$ is a σ-model. Suppose further that $A \subseteq D$, φ is a σ-wff with free variables amongst b, a_1, ..., a_n, and $d_1, ..., d_n \in D$. Then φ* defines A with parameters d_1, ..., d_n *if*

$$A = \{d \in D : \mathcal{M} \models \varphi[d, d_1, ..., d_n]\}$$

Theorem 5.0.11 (Tarski-Vaught test) *Suppose $\mathcal{M}_1$ and $\mathcal{M}_2$ are σ-models, and $D_1 \subseteq D_2$. Then $\mathcal{M}_1$ is an elementary submodel of $\mathcal{M}_2$ if, and only if, for every non-empty subset $A \subseteq D_2$ for which there is a σ-wff that defines A with parameters from D_1, A contains an element of D_1.*

Proof. Suppose $\mathcal{M}_1$ and $\mathcal{M}_2$ are as they are in the hypothesis of the theorem.

('Only if.') Suppose $\mathcal{M}_1$ is an elementary submodel of $\mathcal{M}_2$. That is, $\mathcal{M}_1$ is a submodel of $\mathcal{M}_2$ and, for all σ-wffs φ containing free variables a_1, ..., a_n, and all d_1, ..., d_n in D_1,

$$\mathcal{M}_1 \models \varphi[d_1, ..., d_n] \text{ iff } \mathcal{M}_2 \models \varphi[d_1, ..., d_n]$$

Now suppose that

$$A = \{d \in D_2 : \mathcal{M}_2 \models \varphi[d, d_1, ..., d_n] \text{ for some } \varphi \text{ and } d_1, ..., d_n \in D_1\}$$

and suppose that A is non-empty. Then there is $d \in D_2$ and $d_1, ..., d_n \in D_1$ such that

$$\mathcal{M}_2 \models \varphi[d, d_1, ..., d_n]$$

Therefore,

$$\mathcal{M}_2 \models \exists x \varphi[x, d_1, ..., d_n]$$

(Of course, this is a slight abuse of notation. What we ought to say is that, if ψ is $\exists x\varphi[x/a]$, then $\mathcal{M}_2 \models \psi[d_1, ..., d_n]$, but this is less perspicuous.) Thus, since $\mathcal{M}_1$ is an elementary submodel of $\mathcal{M}_2$,

$$\mathcal{M}_1 \models \exists x\varphi[x, d_1, ..., d_n]$$

So, there is $d' \in D_1$ such that

$$\mathcal{M}_1 \models \varphi[d', d_1, ..., d_n]$$

So, again since $\mathcal{M}_1$ is an elementary submodel of $\mathcal{M}_2$,

$$\mathcal{M}_2 \models \varphi[d', d_1, ..., d_n]$$

Thus, $d' \in A$, as required.

('If.') Now suppose that, for every non-empty subset $A \subseteq D_2$ for which there is a σ-wff that defines A with parameters from D_1, A contains an element of D_1. Then we proceed by induction on the construction of σ-wffs to show that, for any σ-wff φ with free variables a_1, ..., a_n, and any $d_1, ..., d_n \in D_1$

$$\mathcal{M}_1 \models \varphi[d_1, ..., d_n] \text{ iff } \mathcal{M}_2 \models \varphi[d_1, ..., d_n] \tag{5.1}$$

Base Case If φ is an atomic formula, (5.1) holds since $\mathcal{M}_1$ is a submodel of $\mathcal{M}_2$.

Inductive Step If (5.1) holds of φ and ψ, then clearly it holds of $\neg\varphi$ and $\varphi \rightarrow \psi$. The remaining case is $\exists x\varphi[x/a]$. If $\mathcal{M}_1 \models \exists x\varphi[x, d_1, ..., d_n]$, then clearly $\mathcal{M}_2 \models \exists x\varphi[x, d_1, ..., d_n]$ by reasoning similar to that used in the 'Only if.' case above. Thus, suppose

$$\mathcal{M}_2 \models \exists x\varphi[x, d_1, ..., d_n]$$

Then there is $d_0 \in D_2$ such that

$$\mathcal{M}_2 \models \varphi[d_0, d_1, ..., d_n]$$

Thus,

$$d_0 \in A =_{df.} \{d \in D_2 : \mathcal{M}_2 \models \varphi[d_0, d_1, ..., d_n]\}$$

But, by hypothesis, there is $d'_0 \in A \cap D_1$. Thus, by induction hypothesis (5.1),

$$\mathcal{M}_1 \models \varphi[d'_0, d_1, ..., d_n]$$

so

$$\mathcal{M}_1 \models \exists x[x, d_1, ..., d_n]$$

as required.

This completes the induction, and thus the proof of 'If.', and thereby the proof of the theorem. □

With this definition and theorem in hand, we can state and prove the Löwenheim-Skolem theorems.

Theorem 5.0.12 (Löwenheim-Skolem theorems) *Suppose σ is a signature and suppose that $\mathcal{M} = (D, I)$ is an infinite σ-model. Then, if κ is an infinite set and the set of constant and function symbols in σ is less than or equal to κ, then there is an infinite σ-model $\mathcal{M}' = (D', I')$ such that*

(1) $D' \simeq \kappa$ *(that is, D' is the same size as κ); and*

(2) i. *If $\kappa \leq D$, then $\mathcal{M}'$ is an elementary submodel of $\mathcal{M}$*
(*This is known as the* downwards Löwenheim-Skolem theorem.)

ii. *If $D < \kappa$, then $\mathcal{M}$ is an elementary submodel of $\mathcal{M}'$.*
(*This is known as the* upwards Löwenheim-Skolem theorem.)

Proof. Our strategy is quite different in the two cases described by (2)i and (2)ii respectively. We begin with (2)ii, which is the more straightforward, requiring only basic compactness arguments of the sort we encountered in section 5.0.4. The proof of (2)i is given after this; it requires more sophisticated ideas, and appeals to the Tarski-Vaught test from above.

(2)ii Suppose σ and $\mathcal{M}$ are as they are in the hypotheses of the theorem. Suppose that κ is infinite and greater than or equal to the set of constant and function symbols in σ. Finally, suppose that $D < \kappa$.

We begin by extending σ to σ' by adding a set of *new* constant symbols $\{\mathbf{c}_\lambda : \lambda \in \kappa\}$. Then consider the following set of σ'-wffs:

$$S = \mathrm{Th}(\mathcal{M}) \cup \{\mathbf{c}_\lambda \neq \mathbf{c}_\gamma : \lambda, \gamma \in \kappa \text{ and } \lambda \neq \gamma\}$$

Then, since $\mathcal{M}$ is infinite, we can use $\mathcal{M}$ to construct a model for any finite subset of S. Thus, by Compactness, there is a σ'-model $\mathcal{M}' = (D', I')$ of S. It is clear that we can restrict the interpretation function of $\mathcal{M}'$ to give a σ-model $\mathcal{M}'|_\sigma$ of $\mathrm{Th}(\mathcal{M})$. Then the two required facts follow:

(I) $D' \simeq \kappa$

(II) $\mathcal{M}$ is an elementary submodel of $\mathcal{M}'|_\sigma$

To prove (I), note that

$$\kappa \leq D' \leq \kappa + D \leq \kappa + \kappa \simeq \kappa$$

This final equality is an elementary theorem of the arithmetic of infinite sets.

To prove (II), we simply observe that $\mathcal{M}' \models S$.

(2)i Suppose σ and $\mathcal{M} = (D, I)$ are as they are in the hypotheses of the theorem. Suppose that κ is infinite and greater than or equal to the set of constant and function symbols in σ. Finally, suppose that $\kappa \leq D$. Then let A be a subset of D that is the same size as κ. Our strategy is to produce the smallest elementary submodel of $\mathcal{M}$ whose domain contains A. It will turn out that the domain of this minimal elementary submodel is the same size as A, as required.

Suppose φ is a σ-wff whose free variables are among b, a_1, ..., a_n. Then there is a function

$$f_{\varphi(b,a_1,\ldots,a_n)} : D^n \to D$$

such that, for any $d_1, ..., d_n \in D$, exactly one of the following holds:

(a) $\mathcal{M} \models \varphi\left[f_{\varphi(b,a_1,\ldots,a_n)}(d_1, ..., d_n), d_1, ..., d_n\right]$;

(b) $\mathcal{M} \not\models \varphi\,[d, d_1, ..., d_n]$, for all $d \in D$.

Then, given a subset $B \subseteq D$, let

$$F(B) = \{d \in D : \mathcal{M} \models \varphi[d, d_1, ..., d_n] \text{ where } \varphi \text{ is a } \sigma\text{-wff and } d_1, ..., d_n \in B\}$$

That is, $F(B)$ is the union of all subsets of D for which there is a σ-wff that defines that subset with parameters in B. Let

$$F^n(B) = \underbrace{F(F(...F}_{n}(B)...))$$

And let

$$F^\infty(B) = \bigcup_{n=1}^{\infty} F^n(B)$$

Then, given $A \subseteq D$ such that $A \simeq \kappa$, I claim two things:

(I) $F^\infty(A) = \kappa$

(II) The σ-model $\mathcal{M}' = (F^\infty, I)$ is an elementary submodel of $\mathcal{M}$

To prove (I), we use some basic cardinal arithmetic. Suppose $B \subseteq D$ and $\kappa \simeq B$; then

$$F(B) \simeq \bigcup_{n=1}^{\infty} (\{\varphi : \varphi \text{ has } n+1 \text{ free variables}\} \times B^n)$$

and since the set of constant and function symbols in σ is less than or equal to κ, it follows that

$$\kappa \leq F(B) \leq \bigcup_{n=1}^{\infty} \kappa^n \times \kappa^n \simeq \bigcup_{n=1}^{\infty} \kappa \times \kappa \simeq \bigcup_{n=1}^{\infty} \kappa \simeq \aleph_0 \times \kappa \simeq \kappa$$

Thus, it follows that $F(A) \simeq \kappa$, and thus $F(F(A)) \simeq \kappa$, and so on. Then

$$F^\infty(A) = \bigcup_{n=1}^{\infty} F^n(A) \simeq \sum_{n=1}^{\infty} \kappa \simeq \aleph_0 \times \kappa \simeq \kappa$$

as required.

To prove (II), we use the Tarski-Vaught test. Suppose that $C \subseteq D$ is non-empty and suppose that φ defines C with parameters d_1, ..., d_2 from $F^\infty(A)$. We must show that C contains an element of $F^\infty(A)$. Let k be the least number such that $d_1, ..., d_n \in F^k(A)$. Then it is clear that

$$C \subseteq F(F^k(A)) = F^{k+1}(A) \subseteq F^\infty(A)$$

Thus, by the Tarski-Vaught test, $\mathcal{M}' = (F^\infty(A), I)$ is an elementary submodel of $\mathcal{M} = (D, I)$, as required.

This completes the proof of the Löwenheim-Skolem theorems. □

Chapter 6

Modal propositional logic

There is not just one modal propositional logic. Rather, there are many. In this chapter, I describe some of the main ones.

Recall that, in order to describe a logical system, you need to describe three things: its language, its axioms, and its rules of inference. I start by describing the language that all modal propositional logics share (§6.1.1). Then I'll give the axioms and rules of inference for K, the weakest system of modal logic (§6.1.2). Then I'll give the axioms for the stronger theories B, T, $S4$, and $S5$ (§6.1.3). The rules of inference for these stronger systems are the same as for K.

6.1 The syntax

6.1.1 The language shared by all modal propositional logics

The language shared by all modal propositional logics is an extension of the language of ordinary propositional logic.

Thus, there are the following familiar components:

(1) Propositional letters: p, q, r, ... (and with subscripts)

(2) Propositional connectives: $\rightarrow$ and $\neg$.[1]

(3) Punctuation: (and).

Plus the following new component:

[1] As in the propositional and first-order case, it helps to have as few logical symbols as possible. Thus, we choose $\rightarrow$ and $\neg$, and we define $\wedge$, $\vee$, and $\leftrightarrow$ in terms of those.

(4) The necessity operator: $\Box$

And the following grammatical rules tell us how we can string these components together to make wffs. The first three are familiar from classical propositional logic. The final one tells us how we might add the new symbol $\Box$ into our formulae.

(i) Every propositional letter is a wff.

(ii) If φ and ψ are wffs, then $\neg\varphi$ is a wff and $\varphi \to \psi$ is a wff.

(iii) If φ is a wff, then $\Box\varphi$ is a wff.

6.1.2 The axioms and rules of inference for K

K is the minimal formal system for modal propositional logic—that is, it is a subsystem of all others. We present it first.

The axioms of K

The axioms of K belong to two groups:

(A) First: if φ, ψ, and θ are wffs, then following wffs are axioms of K:

- $(\varphi \to (\psi \to \varphi))$
- $((\varphi \to (\psi \to \theta)) \to ((\varphi \to \psi) \to (\varphi \to \theta)))$
- $((\neg\varphi \to \neg\psi) \to (\psi \to \varphi))$

(B) Second: if φ and ψ are wffs, then the following is an axiom of K:

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

Sometimes, this is called the *Axiom of Distributivity*.

The rules of inference of K

The system K has two rules of inference:

(C) The first is the familiar *Rule of Modus Ponens* (or $\to$-elimination) from ordinary propositional logic:

- From $(\varphi \to \psi)$ and φ, infer ψ.

(D) This rule is known as the *Rule of Necessitation*:

- From φ, infer $\Box\varphi$.

  Thus, when you have established φ, you may then infer that φ is necessarily true.

If there is a proof of φ from assumptions in S using these axioms and rules of inference, we write $S \vdash_K \varphi$. Since every wff, axiom, and rule of inference of classical propositional logic also belongs to K, CPL is a subsystem of K—that is, $S \vdash_{CPL} \varphi$ entails $S \vdash_K \varphi$.

6.1.3 The axioms of the stronger systems

We now introduce a variety of systems that are stronger than K. The rules of inference for the stronger systems remain the same as for K. But they all have new axioms over and above those in K. I list them here:

- $T = K + (\Box\varphi \to \varphi)$
- $B = K + (\Box\varphi \to \varphi) + (\neg\varphi \to \Box\neg\Box\varphi)$
- $S4 = K + (\Box\varphi \to \varphi) + (\Box\varphi \to \Box\Box\varphi)$
- $S5 = K + (\Box\varphi \to \varphi) + (\neg\varphi \to \Box\neg\Box\varphi) + (\Box\varphi \to \Box\Box\varphi)$

If L is K, T, B, $S4$, or $S5$, and S is a set of wffs in modal propositional logic, and φ is a wff, then we write

$$S \vdash_L \varphi$$

if one can prove φ from S using only the axioms and rules of inference of L.

This completes my description of various modal propositional logics: we have met K—the weakest system—as well as T, B, $S4$, and $S5$—each of which is stronger than K. Next, we're going to look at a particular sort of *semantics* for these theories that was proposed by Saul Kripke.

6.2 The semantics

6.2.1 The definition of a Kripke model

A *Kripke model* consists of three components:

(1) A set of 'possible worlds', which we call W.

  While we call these 'possible worlds', they need not be worlds in any recognizable sense. They can be anything. For instance, as we will see, they can be sets of wffs.

(2) An *accessibility relation*, which we call R.

This is a relation that may or may not hold between two possible worlds, w_1 and w_2, from W. You might think of w_1Rw_2 as meaning that world w_2 is possible from world w_1.

(3) An *interpretation*, which we call I.

This is a function that takes as its inputs a possible world w and a propositional letter p (for instance), and gives as its output F or T.

- If $I(p, w) = T$, then the propositional letter p is *true at world* w
- If $I(p, w) = F$, then the propositional letter p is *false at world* w.

We denote such a Kripke model as $\mathcal{M} = \langle W, R, I \rangle$: this notation just means that the model is the ordered triple whose first elements is W (the set of possible worlds), whose second element is R (the accessibility relation between worlds), and whose third element is I (the interpretation).

6.2.2 Extending the interpretation to complex wffs

We have said what it means for a propositional letter to be true at a world in a given model, and what it means for a propositional letter to be false at a world in a given model. But what does it mean for a more complex wff to be true or false at a world in a model? We say that now.

At the moment, I is a function that takes only propositional letters and worlds to F or T. Now we show how to extend any such function to a function that takes as its input any wff and any world, and gives as its output F or T. And we abuse notation a bit and use I to denote this extended function as well.

Suppose φ and ψ are wffs and w is a world and suppose that we have defined $I(\varphi, w)$ and $I(\psi, w)$. Then

- $I(\neg\varphi, w) = T$ if $I(\varphi, w) = F$; and $I(\neg\varphi, w) = F$ if $I(\varphi, w) = T$.
- $I((\varphi \to \psi), w) = T$ if $I(\varphi, w) = F$ or $I(\psi, w) = T$; otherwise, $I((\varphi \to \psi), w) = F$
- $I(\Box\varphi, w) = T$ if $I(\varphi, w') = T$ for all $w' \in W$ such that wRw'.

Remarks The first two are the same as in the case of classical propositional logics. However, the third is all new. It says this: $\Box\varphi$ is true at a world if φ is true at all worlds that are accessible from that world. When you remember that w_1Rw_2 means w_2 is possible from w_1, it shouldn't be too difficult to convince yourself that this is the correct definition of when a proposition is necessarily true at a world.

Our interpretation function tells us only when a wff is true *at a world in a Kripke model.* It does not tell us when a wff is true *in a Kripke model.* We define this now:

Definition 6.2.1 (Truth in a model) *Suppose* φ *is a wff and* $\mathcal{M} = \langle W, R, I \rangle$ *is a Kripke model. Then we say that* φ is true in $\mathcal{M}$ (written: $\mathcal{M} \models \varphi$) *if*

For all worlds $w \in W$, $I(\varphi, w) = T$

And we write $\mathcal{M} \models S$ if $\mathcal{M} \models \varphi$ for all φ in S.

6.2.3 Important classes of Kripke models

Certain classes of Kripke models will prove to be important to us when we come to prove the soundness and completeness of our systems of modal propositional logic. I just lay down our abbreviations for them now:

(1) $\mathcal{U}$ is the class of all Kripke models.

(2) $\mathcal{R}$ is the class of all Kripke models in which the accessibility relation is *reflexive*: that is, for all worlds w in W, wRw.

In other words, in a reflexive model, every world is possible from itself.

(3) $\mathcal{S}$ is the class of all Kripke models in which the accessibility relation is *symmetric*: that is, for all worlds w_1 and w_2, if w_1Rw_2, then w_2Rw_1.

In other words, in a symmetric model, if one world is possible from another, that other is possible from the first.

(4) $\mathcal{T}$ is the class of all Kripke models in which the accessibility relation is *transitive*: that is, for all worlds w_1, w_2, and w_3, if w_1Rw_2 and w_2Rw_3, then w_1Rw_3.

In other words, in a transitive model, if one world is possible from another and that other is possible from a third, then the first is possible from the third.

Then we write $\mathcal{RT}$ for the class of models in which the accessibility relation is both reflexive *and* transitive; and we write $\mathcal{ST}$ for the class of models in which the accessibility relation is both symmetric *and* transitive; and so on.

With these definitions in hand, we can define the notions of *logical consequence* and *satisfiability* for modal propositional logic:

Definition 6.2.2 (Logical consequence) *If $\mathcal{C}$ is a class of Kripke models, we write $S \models_{\mathcal{C}} \varphi$ if, for all Kripke models $\mathcal{M}$ in $\mathcal{C}$, if $\mathcal{M} \models S$, then $\mathcal{M} \models \varphi$.*

Definition 6.2.3 *If $\mathcal{C}$ is a class of Kripke models and S is a set of wffs, S* is $\mathcal{C}$-satisfiable *if there is a Kripke model $\mathcal{M}$ in $\mathcal{C}$ such that $\mathcal{M} \models S$.*

6.3 The soundness and completeness theorems

In this section, we ask how the relation of provability in a given system of modal propositional logic is connected to the relation of logical consequence relative to a class of models. To state the connection, we lay down the following definitions:

Definition 6.3.1 (Soundness) *Suppose L is a system of modal propositional logic and suppose $\mathcal{C}$ is a class of Kripke models. Then*

(1) *We say that L* is sound with respect to $\mathcal{C}$ *if*

$$S \vdash_L \varphi \textit{ implies } S \models_{\mathcal{C}} \varphi$$

(2) *We say that L* is complete with respect to $\mathcal{C}$ *if*

$$S \models_{\mathcal{C}} \varphi \textit{ implies } S \vdash_L \varphi$$

The following theorem describes the connections between the relations of provability in a given system of modal propositional logic and the relation of logical consequence relative to a class of models.

Theorem 6.3.2 (Soundness and completeness for K, T, B, $S4$, and $S5$)

(1) *K is sound and complete with respect to the class of all Kripke models.*

(2) *T is sound and complete with respect to $\mathcal{R}$.*

(3) *B is sound and complete with respect to $\mathcal{RS}$.*

(4) *$S4$ is sound and complete with respect to $\mathcal{RT}$.*

(5) *$S5$ is sound and complete with respect to $\mathcal{RST}$.*

To make sure we understand what this means, let's take (4). This means that, φ may be proved from S and the axioms of $S4$ if, and only if, all Kripke models that are both reflexive and transitive and in which each wff in S is true are also models in which φ is true.

6.3.1 The soundness results

Here's the soundness portion of the soundness and completeness theorems to remind us what we're proving:

Theorem 6.3.3 (Soundness for K, T, B, $S4$, and $S5$) *Suppose S is a set of wffs and φ is a wff:*

(1) $S \vdash_K \varphi$ *implies* $S \models_{\mathcal{U}} \varphi$

(2) $S \vdash_T \varphi$ *implies* $S \models_{\mathcal{R}} \varphi$

(3) $S \vdash_B \varphi$ *implies* $S \models_{\mathcal{RS}} \varphi$

(4) $S \vdash_{S4} \varphi$ *implies* $S \models_{\mathcal{RT}} \varphi$

(5) $S \vdash_{S5} \varphi$ *implies* $S \models_{\mathcal{RST}} \varphi$

Here we wish to show that, if φ is provable from S in a given modal propositional logic, then φ is a logical consequence of S relative to the relevant class of models: that is, if S proves φ in a given system, then φ is true in all models in the relevant class in which all wffs in S are true. Thus, it suffices to show two things:

(1) Every axiom of the given modal propositional logic is true in all models in the relevant class.

(2) (a) Given a Kripke model $\mathcal{M}$, the Rule of Modus Ponens preserves the property of being true in $\mathcal{M}$: that is, if $\varphi \to \psi$ and φ are true in $\mathcal{M}$, then ψ is true in $\mathcal{M}$.

(b) Given a Kripke model $\mathcal{M}$, the Rule of Necessitation preserves the property of being true in $\mathcal{M}$: that is, if φ is true in $\mathcal{M}$, then $\Box\varphi$ is true in $\mathcal{M}$.

We prove (2) first, and then prove (1) for each system and the relevant class of models given above:

(2) (a) *Proof.* This is straightforward: if $I(\psi \to \theta, w) = T$ and $I(\psi, w) = T$, then it is clear that $I(\theta, w) = T$ as well.

Thus, the Rule of Modus Ponens preserves the property of being true in $\mathcal{M}$.

(b) *Proof.* Suppose that ψ is true in $\mathcal{M}$: that is, for all worlds w, $I(\psi, w) = T$. Then given a world $w \in W$, it is certainly true that, at all worlds w' accessible from w, $I(\psi, w') = T$, and thus $I(\Box\psi, w) = T$, as required.

Thus, the Rule of Necessitation preserves the property of being true in a model.

Next, we show that

(1) The axioms of K are true in all Kripke models.

(2) The axioms of T are true in all models in $\mathcal{R}$.

(3) The axioms of B are true in all models in $\mathcal{RS}$.

(4) The axioms of $S4$ are true in all models in $\mathcal{RT}$.

(5) The axioms of $S5$ are true in all models in $\mathcal{RST}$.

It suffices to show that

(i) $(\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi))$ is true in all models.

(ii) $(\Box\varphi \to \varphi)$ is true in all reflexive models.

(iii) $(\neg\varphi \to \Box\neg\Box\varphi)$ is true in all symmetric models.

(iv) $(\Box\varphi \to \Box\Box\varphi)$ is true in all transitive models.

(i) *Proof.* Suppose $I(\Box(\varphi \to \psi), w) = T$. Then, for all worlds w' that are accessible from w, $I(\varphi, w') = F$ or $I(\psi, w') = T$. Now, if it is also the case that $I(\Box\varphi, w) = T$ and thus that, for all worlds w' that are accessible from w, $I(\varphi, w') = T$, then it follows that, at all worlds w' that are accessible from w, $I(\psi, w') = T$ and thus that $I(\Box\psi, w) = T$, as required.

(ii) *Proof.* Suppose the accessibility relation is reflexive. And suppose $I(\Box\varphi, w) = T$. Then, since the accessibility relation is *reflexive*, wRw, it follows that $I(\varphi, w) = T$.

(iii) *Proof.* Suppose the accessibility relation is symmetric. And suppose $I(\varphi, w) = F$. And suppose that wRw'. Then it suffices to show that $I(\Box\varphi, w') = F$ (because it will follow that $I(\Box\neg\Box\varphi, w) = T$, as required). Now, since the accessibility relation is *symmetric*, we have $w'Rw$, and thus, since $I(\varphi, w) = F$, it is not true that φ is true at every world accessible from w': in particular, it is not true at w. Thus, $I(\Box\varphi, w') = F$, as required.

(iv) *Proof.* Suppose the accessibility relation is transitive. And suppose $I(\Box\varphi, w) = T$. Then we must show that, for all worlds w such that wRw', $I(\Box\varphi, w') = T$. Suppose w'' is a world and $w'Rw''$. Then, since the accessibility relation is transitive, wRw''. Thus, since $I(\Box\varphi, w) = T$, $I(\varphi, w'') = T$. And thus $I(\Box\varphi, w') = T$. And thus $I(\Box\Box\varphi, w) = T$, as required.

6.3.2 The completeness results

Having proved the soundness portion of the soundness and completeness theorem, we now prove the completeness portion. Here it is to remind us:

Theorem 6.3.4 (Completeness for K, T, B, $S4$, and $S5$) *Suppose S is a set of wffs and φ is a wff:*

(1) $S \models_{\mathcal{U}} \varphi$ *implies* $S \vdash_K \varphi$

(2) $S \models_{\mathcal{R}} \varphi$ *implies* $S \vdash_T \varphi$

(3) $S \models_{\mathcal{RS}} \varphi$ *implies* $S \vdash_B \varphi$

(4) $S \models_{\mathcal{RT}} \varphi$ *implies* $S \vdash_{S4} \varphi$

(5) $S \models_{\mathcal{RST}} \varphi$ *implies* $S \vdash_{S5} \varphi$

To prove this theorem, we prove the following equivalent theorem. To state it, we require the following definition:

Definition 6.3.5 (L-consistent) *Suppose L is a system of modal propositional logic and S is a set of wffs. Then S* is L-consistent *if* $S \nvdash_L \bot$.

Thus, a set of wffs is L-consistent if one cannot derive a contradiction from it in L.

Theorem 6.3.6 (Completeness for K, T, B, $S4$, and $S5$) *Suppose S is a set of wffs:*

(1') *If S is K-consistent, then there is $\mathcal{M}$ in $\mathcal{U}$ such that $\mathcal{M} \models S$.*

(2') *If S is T-consistent, then there is $\mathcal{M}$ in $\mathcal{R}$ such that $\mathcal{M} \models S$.*

(3') *If S is B-consistent, then there is $\mathcal{M}$ in $\mathcal{RS}$ such that $\mathcal{M} \models S$.*

(4') *If S is $S4$-consistent, then there is $\mathcal{M}$ in $\mathcal{RT}$ such that $\mathcal{M} \models S$.*

(5') *If S is $S5$-consistent, then there is $\mathcal{M}$ in $\mathcal{RST}$ such that $\mathcal{M} \models S$.*

The following lemma shows that (1) and (1') are equivalent, (2) and (2') are equivalent, and so on.

Lemma 6.3.7 *Suppose L is a system of modal propositional logic and $\mathcal{C}$ is a class of models. Then the following two propositions are equivalent:*

(1) *For all S and φ, if $S \models_{\mathcal{C}} \varphi$ then $S \vdash_L \varphi$.*

(2) *For all S, if S is L-consistent, then S is $\mathcal{C}$-satisfiable.*

Proof. The proof is exactly similar to that given of Proposition 3.5.6 above. □

Having shown that, in order to prove Theorem 6.3.4, it will suffice to prove Theorem 6.3.6, we now prove Theorem 6.3.6. The key result we need to get that is the following lemma.

Lemma 6.3.8 (The existence of canonical models) *Suppose S is a set of wffs. Then, if S is K-consistent, there is a Kripke model $\mathcal{CM}(S)$ such that $\mathcal{CM}(S) \models S$. We call this model the* canonical model for S.

We devote the next section to proving this, and in the section following, we apply it to prove Theorem 6.3.6.

Constructing canonical models

In this section, we prove that, if S is a K-consistent set of wffs, there is a model, called the *canonical model of S* (written: $\mathcal{CM}(S)$), such that each wff in S is true in $\mathcal{CM}(S)$.

I said above that, while we call the members of W 'possible worlds', they needn't be anything like what we would call a 'possible world' in philosophy. And this is particularly true in the case of canonical models for K-consistent sets of wffs: the 'possible worlds' in a canonical model are all sets of wffs. Bearing this in mind should hopefully make things clearer as we proceed.

More precisely, each world w in the canonical model for S is a maximal and K-consistent set of wffs that includes each member of S. Thus, each world w in our canonical model is the same sort of thing as the models we created for sets of propositional wffs in the chapter on classical propositional logic.

Definition 6.3.9 (Canonical model) *Suppose S is a set of wffs, and suppose that S is K-consistent. Then define $\mathcal{CM}(S) = \langle W, R, v\rangle$ as follows:*

- *The set of worlds W is defined as follows:*

$$W = \{\Gamma : \Gamma \textit{ is maximal and } K\textit{-consistent and } S \subseteq \Gamma\}$$

- *The accessibility relation R is defined as follows: Suppose w_1 and w_2 are in W (therefore, they are sets of wffs!). Then*

$$w_1 R w_2 \textit{ iff, for all wffs } \psi, \textit{ if } \Box\psi \in w_1, \textit{ then } \psi \in w_2.$$

- *The interpretation I is defined as follows: Suppose p is a propositional letter and w is a world (a set of wffs!) in W. Then*

$$I(p, w) = \begin{cases} T & \textit{if } p \in w \\ F & \textit{if } p \notin w \end{cases}$$

Having defined the canonical model $\mathcal{CM}(S)$ for a set of K-consistent wffs S, we must show that it has the property we claimed that it had in Lemma 6.3.8. This is a consequence of the following theorem. The proof of this theorem takes up a lot of space, but it is not difficult. It is simply an induction on the lengths of the wffs.

Theorem 6.3.10 (Fundamental Theorem of Canonical Models) *Suppose S is a K-consistent set of wffs. Then, for any wff ψ and any world w in $\mathcal{CM}(S)$,*

$$\begin{aligned} I(\psi, w) = T \quad \textit{if} \quad \psi \in w \\ I(\psi, w) = F \quad \textit{if} \quad \psi \notin w \end{aligned}$$

Proof. We prove this by *mathematical induction*. That is, first, we prove that it is true for all wffs containing no connectives; we call this the Base Case. Then we show that, for any number n, if it is true for all wffs containing n or fewer connectives, then it is true for all wffs containing $n+1$ connectives; we call this the Inductive Step.

Base Case It is true by the definition of $\mathcal{CM}(S)$ that, for all wffs ψ that contain no connectives—that is, the propositional letters—and for all worlds w

$$\begin{aligned} I(\psi, w) = T \quad \text{if} \quad \psi \in w \\ I(\psi, w) = F \quad \text{if} \quad \psi \notin w \end{aligned}$$

Inductive Step Now, suppose that, for all wffs ψ with n or fewer connectives and for all worlds w

$$\begin{aligned} I(\psi, w) = T \quad &\text{if} \quad \psi \in w \\ I(\psi, w) = F \quad &\text{if} \quad \psi \notin w \end{aligned}$$

We call this the *inductive hypothesis*.

And now suppose that χ is a wff with $n+1$ connectives. Then we must show that for all worlds w

$$\begin{aligned} I(\chi, w) = T \quad &\text{if} \quad \chi \in w \\ I(\chi, w) = F \quad &\text{if} \quad \chi \notin w \end{aligned}$$

There are three cases to consider; the first two are familiar from the case of classical propositional logic:

(1) χ is $\neg\varphi$. Then, since χ contains $n+1$ connectives, φ contains n connectives. Thus, by the inductive hypothesis, for all worlds w

$$\begin{aligned} I(\varphi, w) = T \quad &\text{if} \quad \varphi \in w \\ I(\varphi, w) = F \quad &\text{if} \quad \varphi \notin w \end{aligned}$$

Now, suppose w is world and suppose $\chi \in w$. That is, $\neg\varphi \in w$. Then, since w is K-consistent, $\varphi \notin w$. Thus, $I(\varphi, w) = F$, so $I(\neg\varphi, w) = T$, as required. Similarly, if $\chi \notin w$, then $\varphi \in w$, so $I(\varphi, w) = T$, so $I(\neg\varphi, w) = F$, as required.

(2) χ is $\varphi \to \psi$. Then, since χ contains $n+1$ connectives, φ and ψ contain n or fewer connectives. Thus, by the inductive hypothesis, for all worlds w,

$$\begin{aligned} I(\varphi, w) = T \quad &\text{if} \quad \varphi \in w \qquad & I(\psi, w) = T \quad &\text{if} \quad \psi \in w \\ I(\varphi, w) = F \quad &\text{if} \quad \varphi \notin w \qquad & I(\psi, w) = F \quad &\text{if} \quad \psi \notin w \end{aligned}$$

Then suppose $\chi \in w$. Then it is not the case φ and $\neg\psi$ are both in w. Thus, either $\neg\varphi \in w$, in which case $I(\varphi, w) = F$ and thus $I(\chi, w) = T$, or $\psi \in w$, in which case $I(\psi, w) = T$ and thus $I(\chi, w) = T$, as required. Similarly, if $\chi \notin w$, then $\neg\chi \in w$ and thus φ and $\neg\varphi$ are in w, and so $I(\varphi, w) = T$ and $I(\psi, w) = F$ and thus $I(\chi, w) = F$.

(3) χ is $\Box\varphi$. Then, since χ contains $n+1$ connectives, φ contains n connectives. Thus, by the inductive hypothesis, for all worlds w

$$I(\varphi, w) = T \quad \text{if} \quad \varphi \in w$$
$$I(\varphi, w) = F \quad \text{if} \quad \varphi \notin w$$

Now suppose $\chi = \Box\varphi \in w$. We must show that, for all worlds $w' \in W$ such that wRw', $\varphi \in w$. This follows immediately from the definition of R in $\mathcal{CM}(S)$. Recall that $w_1 R w_2$ iff, for every wff ψ, if $\Box\psi \in w_1$, then $\psi \in w_2$. Thus, if wRw' and $\Box\varphi \in w$, then $\varphi \in w'$, so by inductive hypothesis $I(\varphi, w') = T$. Thus, $I(\Box\varphi, w) = T$.

It is considerably more difficult to prove that, if $\Box\varphi \notin w$, then $I(\Box\varphi, w) = F$. We omit the proof here.

This completes our proof. □

In the next section, we apply this result to prove Theorem 6.3.6.

Proof of Theorem 6.3.6

We will prove each component of Theorem 6.3.6 in turn.

(1') *Proof.* This is the content of Lemma 6.3.8. If S is K-consistent, $\mathcal{CM}(S)$ is a Kripke model and $\mathcal{CM}(S) \models S$, as required.

(2') *Proof.* Suppose S is T-consistent. Then, since

$$T = K + (\Box\varphi \to \varphi)$$

it follows that $S \cup \{\Box\varphi \to \varphi\}$ is K-consistent. Thus, by the Fundamental Theorem of Canonical Models

$$\mathcal{CM}(S \cup \{\Box\varphi \to \varphi\}) \models S \cup \{\Box\varphi \to \varphi\}$$

Thus, in particular

$$\mathcal{CM}(S \cup \{\Box\varphi \to \varphi\}) \models S$$

Thus, to complete our proof, we need only show that the accessibility relation in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi\})$ is reflexive.

By the Fundamental Theorem of Canonical Models, $(\Box\varphi \to \varphi)$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi\})$. We use this to show that, in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi\})$, the accessibility relation in reflexive.

Suppose w is a world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi\})$. Then we must show that wRw. By the definition of R in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi\})$, this requires us to show that, for any wff ψ, if $\Box\psi \in w$, then $\psi \in w$. This is clearly true, since $(\Box\psi \to \psi$ is true at w.

(3') *Proof.* Suppose S is B-consistent. Then, since

$$B = K + (\Box\varphi \to \varphi) + (\neg\varphi \to \Box\neg\Box\varphi)$$

it follows that $S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\}$ is K-consistent. Thus, by the Fundamental Theorem of Canonical Models

$$\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\}) \models S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\}$$

Thus, in particular

$$\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\}) \models S$$

Thus, to complete our proof, we need only show that the accessibility relation in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\})$ is both reflexive and symmetric.

By the Fundamental Theorem of Canonical Models, $(\Box\varphi \to \varphi)$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\})$ and from our proof of (2'), we know that it follows from this that the accessibility relation is reflexive.

Also, by the Fundamental Theorem of Canonical Models, $\neg\varphi \to \Box\neg\Box\varphi$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\})$. We use this to show that, in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\})$, the accessibility relation in symmetric as well.

Suppose w_1 and w_2 are worlds in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi\})$. And suppose that w_1Rw_2. Thus, for any wff ψ, if $\Box\psi \in w_1$, then $\psi \in w_2$. Now, we must show that w_2Rw_1. That is, we must show that, for any wff ψ, if $\Box\psi \in w_2$, then $\psi \in w_1$. Suppose $\Box\psi \in w_2$. And suppose (for a contradiction) that $\psi \notin w_1$. Then, since each world is maximally K-consistent, $\neg\psi \in w_1$. Thus, since $\neg\psi \to \Box\neg\Box\psi$ is true at w_1, it follows that $\Box\neg\Box\psi$ is true at w_1. Thus, since w_1Rw_2, it follows that $\neg\Box\varphi$ is in w_2; but this gives a contradiction, since we have assumed that $\Box\varphi$ is in w_2 and that w_2 is K-consistent. Thus, $\psi \in w_1$, as required, and so w_2Rw_1.

(4') Suppose S is $S4$-consistent. Then, since

$$S4 = K + (\Box\varphi \to \varphi) + (\Box\varphi \to \Box\Box\varphi)$$

it follows that $S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\}$ is K-consistent. Thus, by the Fundamental Theorem of Canonical Models

$$\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\}) \models S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\}$$

Thus, in particular

$$\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\}) \models S$$

Thus, to complete our proof, we need only show that the accessibility relation in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\})$ is both reflexive and transitive.

By the Fundamental Theorem of Canonical Models, $\Box\varphi \to \varphi$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\})$ and from our proof of (2'), we know that it follows from this that the accessibility relation is reflexive.

Also, by the Fundamental Theorem of Canonical Models, $\Box\varphi \to \Box\Box\varphi$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\})$. We use this to show that, in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\})$, the accessibility relation in transitive as well.

Suppose w_1, w_2, and w_3 are worlds in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi\})$. And suppose that

(i) $w_1 R w_2$: that is, for all ψ, if $\Box\psi \in w_1$, then $\psi \in w_2$; and

(ii) $w_2 R w_3$: that is, for all ψ, if $\Box\psi \in w_2$, then $\psi \in w_3$

We must show that $w_1 R w_3$: that is, we must show that, for all ψ, if $\Box\psi \in w_1$, then $\psi \in w_3$. Thus, suppose $\Box\psi \in w_1$. Then, since $\Box\psi \to \Box\Box\psi \in w_1$, it follows that $\Box\Box\psi \in w_1$. Then, by (i), it follows that $\Box\psi \in w_2$. And, by (ii), it follows that $\psi \in w_3$, as required.

(5') Suppose S is $S5$-consistent. Then, since

$$S4 = K + (\Box\varphi \to \varphi) + (\neg\varphi \to \Box\neg\Box\varphi) + (\Box\varphi \to \Box\Box\varphi)$$

it follows that $S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi, \Box\varphi \to \Box\Box\varphi\}$ is K-consistent. Thus, by the Fundamental Theorem of Canonical Models

$$\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi, \Box\varphi \to \Box\Box\varphi\}) \models S$$

Thus, to complete our proof, we need only show that the accessibility relation in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi, \Box\varphi \to \Box\Box\varphi\})$ is reflexive, symmetric, *and* transitive.

By our proof of (2'), since $\Box\varphi \to \varphi$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi, \Box\varphi \to \Box\Box\varphi\})$ the accessibility relation is reflexive.

By our proof of (3'), since $\neg\varphi \to \Box\neg\Box\varphi$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi, \Box\varphi \to \Box\Box\varphi\})$ the accessibility relation is symmetric.

By our proof of (4'), since $\Box\varphi \to \Box\Box\varphi$ is true at every world in $\mathcal{CM}(S \cup \{\Box\varphi \to \varphi, \neg\varphi \to \Box\neg\Box\varphi, \Box\varphi \to \Box\Box\varphi\})$ the accessibility relation is transitive.

This completes our proof of the completeness portion of the soundness and completeness theorems. Thus, we have proved the soundness and completeness theorems for K, B, T, $S4$, and $S5$. $\Box$